

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/49, A61K 48/00		(11) International Publication Number: WO 00/39302
		(43) International Publication Date: 6 July 2000 (06.07.00)
<p>(21) International Application Number: PCT/US99/31245</p> <p>(22) International Filing Date: 30 December 1999 (30.12.99)</p> <p>(30) Priority Data: 60/114,495 31 December 1998 (31.12.98) US 60/168,471 1 December 1999 (01.12.99) US</p> <p>(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).</p>		
<p>(72) Inventors: BARNETT, Susan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). ZUR MEGEDE, Jan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). SRIVASTAVA, Indresh; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). LIAN, Ying; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). HARTOG, Karin; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). LIU, Hong; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). GREER, Catherine; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). SELBY, Mark; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). WALKER, Christopher; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US).</p> <p>(74) Agents: DOLLARD, Anne, S.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US) et al.</p> <p>(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published Without international search report and to be republished upon receipt of that report.</p>		

(54) Title: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION OF VIRUS-LIKE PARTICLES

(57) Abstract

The present invention relates to the efficient expression of HIV polypeptides in a variety of cell types, including, but not limited to, mammalian, insect, and plant cells. Synthetic expression cassettes encoding the HIV Gag-containing polypeptides are described, as are uses of the expression cassettes in applications including DNA immunization, generation of packaging cell lines, and production of Env-, tat- or Gag-containing proteins. The invention provides methods of producing Virus-Like Particles (VLPs), as well as, uses of the VLPs including, but not limited to, vehicles for the presentation of antigens and stimulation of immune response in subjects to whom the VLPs are administered.

BEST AVAILABLE COPY

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND
PRODUCTION OF VIRUS-LIKE PARTICLES

5 TECHNICAL FIELD

Synthetic expression cassettes encoding the HIV polypeptides (e.g., Gag-, pol-, prot-, reverse transcriptase, Env- or tat-containing polypeptides) are described, as are uses of the expression cassettes. The present invention relates to the efficient expression of HIV polypeptides in a variety of cell types. Further, the invention provides methods of producing Virus-Like Particles (VLPs), as well as, uses of the VLPs and high level expression of oligomeric envelope proteins.

15

BACKGROUND OF THE INVENTION

Acquired immune deficiency syndrome (AIDS) is recognized as one of the greatest health threats facing modern medicine. There is, as yet, no cure for this disease.

In 1983-1984, three groups independently identified the suspected etiological agent of AIDS. See, e.g., Barre-Sinoussi et al. (1983) Science 220:868-871; Montagnier et al., in Human T-Cell Leukemia Viruses (Gallo, Essex & Gross, eds., 1984); Vilmer et al. (1984) The Lancet 1:753; Popovic et al. (1984) Science 224:497-500; Levy et al. (1984) Science 225:840-842. These isolates were variously called lymphadenopathy-associated virus (LAV), human T-cell lymphotropic virus

type III (HTLV-III), or AIDS-associated retrovirus (ARV). All of these isolates are strains of the same virus, and were later collectively named Human Immunodeficiency Virus (HIV). With the isolation of a related

5 AIDS-causing virus, the strains originally called HIV are now termed HIV-1 and the related virus is called HIV-2. See, e.g., Guyader et al. (1987) *Nature* 326:662-669; Brun-Vezinet et al. (1986) *Science* 233:343-346; Clavel et al. (1986) *Nature* 324:691-695.

10 A great deal of information has been gathered about the HIV virus, however, to date an effective vaccine has not been identified. Several targets for vaccine development have been examined including the env, Gag, pol and tat gene products encoded by HIV.

15 Haas, et al., (*Current Biology* 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (*J. Virol.* 72(2):1497-1503, 1998) described an increased immune response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage. Schneider, et al., (*J. Virol.* 71(7):4892-4903, 1997) discuss inactivation of inhibitory (or instability) elements (INS) located within the coding sequences of the 20 Gag and Gag-protease coding sequences.

25 The Gag proteins of HIV-1 are necessary for the assembly of virus-like particles. HIV-1 Gag proteins are involved in many stages of the life cycle of the virus including, assembly, virion maturation after particle release, and early post-entry steps in virus replication. The roles of HIV-1 Gag proteins are numerous and complex (Freed, E.O., *Virology* 251:1-15, 1998).

Wolf, et al., (PCT International Application, WO 96/30523, published 3 October 1996; European Patent Application, Publication No. 0 449 116 A1, published 2 October 1991) have described the use of altered pr55 Gag of HIV-1 to act as a non-infectious retroviral-like particulate carrier, in particular, for the presentation of immunologically important epitopes. Wang, et al., (Virology 200:524-534, 1994) describe a system to study assembly of HIV Gag- β -galactosidase fusion proteins into virions. They describe the construction of sequences encoding HIV Gag- β -galactosidase fusion proteins, the expression of such sequences in the presence of HIV Gag proteins, and assembly of these proteins into virus particles.

Recently, Shiver, et al., (PCT International Application, WO 98/34640, published 13 August 1998) described altering HIV-1 (CAM1) Gag coding sequences to produce synthetic DNA molecules encoding HIV Gag and modifications of HIV Gag. The codons of the synthetic molecules were codons preferred by a projected host cell.

The envelope protein of HIV-1 is a glycoprotein of about 160 kD (gp160). During virus infection of the host cell, gp160 is cleaved by host cell proteases to form gp120 and the integral membrane protein, gp41. The gp41 portion is anchored in (and spans) the membrane bilayer of virion, while the gp120 segment protrudes into the surrounding environment. As there is no covalent attachment between gp120 and gp41, free gp120 is released from the surface of virions and infected cells.

Haas, et al., (Current Biology 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (J. Virol.

72(2):1497-1503, 1998) described an increased immune response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage.

5 **SUMMARY OF THE INVENTION**

The present invention relates to improved expression of HIV Env-, tat-, pol-, prot-, reverse transcriptase, or Gag-containing polypeptides and production of virus-like particles.

10 In one embodiment the present invention includes an expression cassette, comprising a polynucleotide encoding an HIV Gag polypeptide comprising a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20. In certain embodiments, the polynucleotide sequence encoding said Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9 or SEQ ID NO:4. The expression cassettes may further include a polynucleotide sequence encoding an HIV protease polypeptide, for example a nucleotide sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:78, and SEQ ID NO:79. The expression cassettes may further include a polynucleotide sequence encoding an HIV reverse transcriptase polypeptide, for example a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, and SEQ ID NO:84. The expression cassettes may further include a polynucleotide sequence encoding an HIV tat polypeptide, for example a sequence selected from the group consisting of: SEQ ID NO:87, SEQ ID NO:88, and SEQ ID NO:89. The expression cassettes may further include a polynucleotide sequence encoding an HIV polymerase polypeptide, for example a

sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6. The expression cassettes may include a polynucleotide sequence encoding an HIV polymerase polypeptide, wherein (i) the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:4, and (ii) wherein the sequence is modified by deletions of coding regions corresponding to reverse transcriptase and integrase. The expression cassettes described above may preserves T-helper cell and CTL epitopes. The expression cassettes may further include a polynucleotide sequence encoding an HCV core polypeptide, for example a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:7.

In another aspect, the invention includes an expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59). In certain embodiments, the Env expression cassettes includes sequences flanking a V1 region but have a deletion in the V1 region itself, for example the sequence presented as SEQ ID NO:65 (Figure 52, gp160.modUS4.delV1). In certain embodiments, the Env expression cassettes, include sequences flanking a V2 region but have a deletion in the V2 region itself, for example the sequences shown in SEQ ID NO:60 (Figure 47); SEQ ID NO:66 (Figure 53); SEQ ID NO:34 (Figure 20); SEQ ID NO:37 (Figure 24); SEQ ID NO:40 (Figure 27); SEQ ID NO:43 (Figure 30); SEQ ID NO:46 (Figure 33); SEQ ID NO:76 (Figure 64) and SEQ ID NO:49 (Figure 36). In certain

embodiments, the Env expression cassettes include sequences flanking a V1/V2 region but have a deletion in the V1/V2 region itself, for example, SEQ ID NO:59 (Figure 46); SEQ ID NO:61 (Figure 48); SEQ ID NO:67 (Figure 54); SEQ ID NO:75 (Figure 63); SEQ ID NO:35 (Figure 21); SEQ ID NO:38 (Figure 25); SEQ ID NO:41 (Figure 28); SEQ ID NO:44 (Figure 31); SEQ ID NO:47 (Figure 34) and SEQ ID NO:50 (Figure 37). The Env-encoding expression cassettes may also include a mutated cleavage site that prevents the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide, for example, SEQ ID NO:57 (Figure 44); SEQ ID NO:61 (Figure 48); SEQ ID NO:63 (Figure 50); SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34). The Env expression cassettes may include a gp160 Env polypeptide or a polypeptide derived from a gp160 Env polypeptide, for example SEQ ID NO:64 (Figure 51); SEQ ID NO:65 (Figure 52); SEQ ID NO:66 (Figure 53); SEQ ID NO:67 (Figure 54); SEQ ID NO:68 (Figure 55); SEQ ID NO:75 (Figure 63); SEQ ID NO:73 (Figure 61); SEQ ID NO:48 (Figure 35); SEQ ID NO:49 (Figure 36); SEQ ID NO:50 (Figure 37); SEQ ID NO:76 (Figure 64); and SEQ ID NO:74 (Figure 62). The Env expression cassettes may include a gp140 Env polypeptide or a polypeptide derived from a gp140 Env polypeptide, for example SEQ ID NO:56 (Figure 43); SEQ ID NO:57 (Figure 44); SEQ ID NO:58 (Figure 45); SEQ ID NO:59 (Figure 46); SEQ ID NO:60 (Figure 47); SEQ ID NO:61 (Figure 48); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); SEQ ID NO:36 (Figure 23); SEQ ID NO:37 (Figure 24); SEQ ID NO:38 (Figure 25); SEQ ID NO:39

(Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34). The Env expression cassettes may also include a gp120 Env polypeptide or a polypeptide derived from a gp120 Env polypeptide, for example SEQ ID NO:54 (Figure 41); and SEQ ID NO:55 (Figure 42); SEQ ID NO:33 (Figure 19); SEQ ID NO:34 (Figure 20); and SEQ ID NO:35 (Figure 21). The Env expression cassettes may include an Env polypeptide lacking the amino acids corresponding to residues 128 to about 194, relative to strains SF162 or US4, for example, SEQ ID NO:55 (Figure 42); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); and SEQ ID NO:68 (Figure 55).

In another aspect, the invention includes a recombinant expression system for use in a selected host cell, comprising, one or more of the expression cassettes described herein operably linked to control elements compatible with expression in the selected host cell. The expression cassettes may be included on one or on multiple vectors and may use the same or different promoters. Exemplary control elements include a transcription promoter (e.g., CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein), a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

In another aspect, the invention includes a recombinant expression system for use in a selected host cell, comprising, any one of the expression cassettes described herein operably linked to control elements

compatible with expression in the selected host cell. Exemplary control elements include, but are not limited to, a transcription promoter (e.g., CMV, CMV+intron A, SV40, RSV, HIV-LTR, MMLV-LTR, and metallothionein), a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

In yet another aspect, the invention includes a cell comprising one or more of the expression cassettes described herein operably linked to control elements compatible with expression in the cell. The cell can be, for example, a mammalian cell (e.g., BHK, VERO, HT1080, 293, RD, COS-7, or CHO cells), an insect cell (e.g., *Trichoplusia ni* (Tn5) or Sf9), a bacterial cell, a plant cell, a yeast cell, an antigen presenting cell (e.g., primary, immortalized or tumor-derived lymphoid cells such as macrophages, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof).

In another aspect, the invention includes methods for producing a polypeptide including HIV Gag-, prot-, pol-, reverse transcriptase, Env- or Tat-containing polypeptide sequences, said method comprising, incubating the cells comprising one or more the expression cassettes described herein, under conditions for producing said polypeptide.

In yet another aspect, the invention includes compositions for generating an immunological response, comprising one or more of the expression cassettes described herein. In certain embodiments, the compositions also include an adjuvant.

In a still further aspect, the invention includes methods of generating an immune response in a subject, comprising introducing a composition comprising one or

more of the expression cassettes described herein into the subject under conditions that are compatible with expression of said expression cassette in the subject. In certain embodiments, the expression cassette is 5 introduced using a gene delivery vector. More than one expression cassette may be introduced using one or more gene delivery vectors.

In yet another aspect, the invention includes a purified polynucleotide comprising a polynucleotide sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59). Further exemplary purified 15 polynucleotide sequences were presented above.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

In another embodiment, the invention includes a 20 method for producing a polypeptide including HIV Gag polypeptide sequences, where the method comprises incubating any of the above cells containing an expression cassette of interest under conditions for producing the polypeptide.

25 The invention further includes, a method for producing virus-like particles (VLPs) where the method comprises incubating any of the above-described cells containing an expression cassette of interest under conditions for producing VLPs.

30 In another aspect the invention includes a method for producing a composition of virus-like particles (VLPs) where, any of the above-described cells containing an expression cassette of interest are incubated under

conditions for producing VLPs, and the VLPs are substantially purified to produce a composition of VLPs.

In a further embodiment of the present invention, packaging cell lines are produced using the expression cassettes of the present invention. For example, a cell line useful for packaging lentivirus vectors comprises suitable host cells that have an expression vector containing an expression cassette of the present invention wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the host cell. In a preferred embodiment, such host cells may be transfected with one or more expression cassettes having a polynucleotide sequence that encodes an HIV polymerase polypeptide or polypeptides derived therefrom, for example, where the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6. Further, the HIV polymerase polypeptide may be modified by deletions of coding regions corresponding to reverse transcriptase and integrase. Such a polynucleotide sequence may preserve T-helper cell and CTL epitopes, for example when used in a vaccine application. In addition, the polynucleotide sequence may also include other polypeptides. Further, polynucleotide sequences encoding additional polypeptides whose expression are useful for packaging cell line function may also be utilized.

In another aspect, the present invention includes a gene delivery or vaccine vector for use in a subject, where the vector is a suitable gene delivery vector for use in the subject, and the vector comprises one or more of any of the expression cassettes of the present

invention where the polynucleotide sequences of interest are operably linked to control elements compatible with expression in the subject. Such gene delivery vectors can be used in a method of DNA immunization of a subject, 5 for example, by introducing a gene delivery vector into the subject under conditions that are compatible with expression of the expression cassette in the subject. Gene delivery vectors useful in the practice of the present invention include, but are not limited to, 10 nonviral vectors, bacterial plasmid vectors, viral vectors, particulate carriers (where the vector is coated on a polylactide co-glycolide particles, gold or tungsten particle, for example, the coated particle can be delivered to a subject cell using a gene gun), liposome preparations, and viral vectors (e.g., vectors derived 15 from alphaviruses, pox viruses, and vaccinia viruses, as well as, retroviral vectors, including, but not limited to, lentiviral vectors). Alphavirus-derived vectors include, for example, an alphavirus cDNA construct, a recombinant alphavirus particle preparation and a eukaryotic layered vector initiation system. In one embodiment, the subject is a vertebrate, preferably a mammal, and in a further embodiment the subject is a human.

25 The invention further includes a method of generating an immune response in a subject, where cells of a subject are transfected with any of the above-described gene delivery vectors (e.g., alphavirus constructs; alphavirus cDNA constructs; eukaryotic layered vector initiation systems (see, e.g., U.S. Patent 30 Number 5,814,482 for description of suitable eukaryotic layered vector initiation systems); alphavirus particle

preparations; etc.) under conditions that permit the expression of a selected polynucleotide and production of a polypeptide of interest (i.e., encoded by any expression cassette of the present invention), thereby eliciting an immunological response to the polypeptide. Transfection of the cells may be performed *ex vivo* and the transfected cells are reintroduced into the subject. Alternately, or in addition, the cells may be transfected *in vivo* in the subject. The immune response may be 10 humoral and/or cell-mediated (cellular).

Further embodiments of the present invention include purified polynucleotides. In one embodiment, the purified polynucleotide comprises a polynucleotide sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20, and complements thereof. In another embodiment, the purified polynucleotide comprises a polynucleotide sequence encoding an HIV Gag polypeptide, wherein the polynucleotide sequence comprises a sequence having at 20 least 90% sequence identity to the sequence presented as SEQ ID NO:20, and complements thereof. In still another embodiment, the purified polynucleotide comprises a polynucleotide sequence encoding an HIV Gag polypeptide, wherein the polynucleotide sequence comprises a sequence 25 having at least 90% sequence identity to the sequence presented as SEQ ID NO:9, and complements thereof. In further embodiments the polynucleotide sequence comprises a sequence having at least 90% sequence identity to one of the following sequences: SEQ ID NO:4, SEQ ID NO:5, SEQ 30 ID NO:6, SEQ ID NO:7, and complements thereof.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

These and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

5 BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the locations of the inactivation sites for the native HIV-1SF2 Gag protein coding sequence.

10 Figure 2 shows the locations of the inactivation sites for the native HIV-1SF2 Gag-protease protein coding sequence.

Figures 3A and 3B show electron micrographs of virus-like particles. Figure 3A shows immature p55Gag virus-like particles in COS-7 cells transfected with a 15 synthetic HIV-1_{SF2} gag construct while Figure 3B shows mature (arrows) and immature VLP in cells transfected with a modified HIV-1_{SF2} gagprotease construct (GP2, SEQ ID NO:70). Transfected cells were fixed at 24 h (gag) or 48 h (gagprotease) post-transfection and subsequently 20 analyzed by electron microscopy (magnification at 100,000X). Cells transfected with vector alone (pCMVKm2) served as negative control (data not shown).

Figure 4 presents an image of samples from a series 25 of fractions which were electrophoresed on an 8-16% SDS polyacrylamide gel and the resulting bands visualized by commassie blue staining. The results show that the native p55 Gag virus-like particles (VLPs) banded at a sucrose density of range of 1.15 - 1.19 g/ml with the peak at approximately 1.17 g/ml.

30 Figure 5 presents an image similar to Figure 4 where the analysis was performed using Gag VLPs produced by a synthetic Gag expression cassette.

Figure 6 presents a comparison of the total amount of purified HIV p55 Gag from several preparations obtained from two baculovirus expression cassettes encoding native and modified Gag.

5 Figure 7 presents an alignment of modified coding sequences of the present invention including a synthetic Gag expression cassette (SEQ ID NO:4), a synthetic Gag-protease expression cassette (SEQ ID NO:5), and a synthetic Gag-polymerase expression cassette (SEQ ID NO:6). A common region (Gag-common; SEQ ID NO:9) extends 10 from position 1 to position 1262.

15 Figure 8 presents an image of wild-type Gag-HCV core expression samples from a series of fractions which were electrophoresed on an 8-16% SDS polyacrylamide gel and the resulting bands visualized by commassie staining.

Figure 9 shows the results of Western blot analysis of the gel shown presented in Figure 8.

20 Figure 10 presents results similar to those shown in Figure 9. The results in Figure 10 indicate that the main HCV Core-specific reactivity migrates at an approximate molecular weight of 72,000 kD, which is in accordance with the predicted molecular weight of the Gag-HCV core chimeric protein.

25 Figures 11A to 11D present a comparison of AT content, in percent, of cDNAs corresponding to an unstable human mRNA (human IFN γ mRNA; 11A), wild-type HIV Gag native RNA (11B), a stable human mRNA (human GAPDH mRNA; 11C), and synthetic HIV Gag RNA (11D).

30 Figure 12 shows the location of the inactivation sites for the native HIV-1SF2 Gag-polymerase sequence.

Figure 13A presents a vector map of pESN2dhfr.

Figure 13B presents a map of the pCMVIII vector.

Figure 14 presents a vector map of pCMV-LINK.

Figure 15 presents a schematic diagram showing the relationships between the following forms of the HIV Env polypeptide: gp160, gp140, gp120, and gp41.

5 Figure 16 depicts the nucleotide sequence of wild-type gp120 from SF162 (SEQ ID NO:30).

Figure 17 depicts the nucleotide sequence of the wild-type gp140 from SF162 (SEQ ID NO:31).

Figure 18 depicts the nucleotide sequence of the wild-type gp160 from SF162 (SEQ ID NO:32).

10 Figure 19 depicts the nucleotide sequence of the construct designated gp120.modSF162 (SEQ ID NO:33).

Figure 20 depicts the nucleotide sequence of the construct designated gp120.modSF162.delV2 (SEQ ID NO:34).

15 Figure 21 depicts the nucleotide sequence of the construct designated gp120.modSF162.delV1/V2 (SEQ ID NO:35).

Figures 22A-H show the percent A-T content over the length of the sequences for IFN γ (Figures 2C and 2G); native gp160 Env US4 and SF162 (Figures 2A and 2E, respectively); GAPDH (Figures 2D and 2H); and the synthetic gp160 Env for US4 and SF162 (Figures 2B and 2F, respectively).

Figure 23 depicts the nucleotide sequence of the construct designated gp140.modSF162 (SEQ ID NO:36).

25 Figure 24 depicts the nucleotide sequence of the construct designated gp140.modSF162.delV2 (SEQ ID NO:37).

Figure 25 depicts the nucleotide sequence of the construct designated gp140.modSF162.delV1/V2 (SEQ ID NO:38).

30 Figure 26 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162 (SEQ ID NO:39).

Figure 27 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162.delV2 (SEQ ID NO:40).

Figure 28 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162.delV1/V2 (SEQ ID NO:41).

5 Figure 29 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162 (SEQ ID NO:42).

Figure 30 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162.delV2 (SEQ ID NO:43).

10 Figure 31 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162.delV1/V2 (SEQ ID NO:44).

Figure 32 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162 (SEQ ID NO:45).

15 Figure 33 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162.delV2 (SEQ ID NO:46).

Figure 34 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162.delV1/V2 (SEQ ID NO:47).

20 Figure 35 depicts the nucleotide sequence of the construct designated gp160.modSF162 (SEQ ID NO:48).

Figure 36 depicts the nucleotide sequence of the construct designated gp160.modSF162.delV2 (SEQ ID NO:49).

25 Figure 37 depicts the nucleotide sequence of the construct designated gp160.modSF162.delV1/V2 (SEQ ID NO:50).

Figure 38 depicts the nucleotide sequence of the wild-type gp120 from US4 (SEQ ID NO:51).

30 Figure 39 depicts the nucleotide sequence of the wild-type gp140 from US4 (SEQ ID NO:52).

Figure 40 depicts the nucleotide sequence of the wild-type gp160 from US4 (SEQ ID NO:53).

Figure 41 depicts the nucleotide sequence of the construct designated gp120.modUS4 (SEQ ID NO:54).

Figure 42 depicts the nucleotide sequence of the construct designated gp120.modUS4.del 128-194 (SEQ ID NO:55).

5 Figure 43 depicts the nucleotide sequence of the construct designated gp140.modUS4 (SEQ ID NO:56).

Figure 44 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4 (SEQ ID NO:57).

Figure 45 depicts the nucleotide sequence of the construct designated gp140.TM.modUS4 (SEQ ID NO:58).

10 Figure 46 depicts the nucleotide sequence of the construct designated gp140.modUS4.delV1/V2 (SEQ ID NO:59).

Figure 47 depicts the nucleotide sequence of the construct designated gp140.modUS4.delV2 (SEQ ID NO:60).

15 Figure 48 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4.delV1/V2 (SEQ ID NO:61).

20 Figure 49 depicts the nucleotide sequence of the construct designated gp140.modUS4.del 128-194 (SEQ ID NO:62).

Figure 50 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4.del 128-194 (SEQ ID NO:63).

25 Figure 51 depicts the nucleotide sequence of the construct designated gp160.modUS4 (SEQ ID NO:64).

Figure 52 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV1 (SEQ ID NO:65).

Figure 53 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV2 (SEQ ID NO:66).

30 Figure 54 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV1/V2 (SEQ ID NO:67).

Figure 55 depicts the nucleotide sequence of the construct designated gp160.modUS4.del 128-194 (SEQ ID NO:68).

5 Figure 56 depicts the nucleotide sequence of the common region of Env from wild-type US4 (SEQ ID NO:69).

Figure 57 depicts the nucleotide sequence of the common region of Env from wild-type SF162 (SEQ ID NO:70).

10 Figure 58 depicts the nucleotide sequence of synthetic sequences corresponding to the common region of Env from US4 (SEQ ID NO:71).

Figure 59 depicts the nucleotide sequence of synthetic sequences corresponding to the common region of Env from SF162 (SEQ ID NO:72).

15 Figure 60 presents a schematic representation of an Env polypeptide purification strategy.

Figure 61 depicts the nucleotide sequence of the bicistronic construct designated gp160.modUS4.Gag.modSF2 (SEQ ID NO:73).

20 Figure 62 depicts the nucleotide sequence of the bicistronic construct designated gp160.modSF162.Gag.modSF2 (SEQ ID NO:74).

Figure 63 depicts the nucleotide sequence of the bicistronic construct designated gp160.modUS4.-delV1/V2.Gag.modSF2 (SEQ ID NO:75).

25 Figure 64 depicts the nucleotide sequence of the bicistronic construct designated gp160.modSF162.delV2.Gag.modSF2 (SEQ ID NO:76).

30 Figures 65A-65F show micrographs of 293T cells transfected with the following polypeptide encoding sequences: Figure 65A, gag.modSF2; Figure 65B, gp160.modUS4; Figure 65C, gp160.modUS4.delV1/V2.gag.modSF2 (bicistronic Env and Gag); Figures 65D and 65E, gp160.modUS4.delV1/V2 and

gag.modSF2; and Figure 65F, gp120.modSF162.delV2 and gag.modSF2.

Figures 66A and 66B present alignments of selected modified coding sequences of the present invention including a common region defined for each group of synthetic Env expression cassettes. Figure 66A presents alignments of modified SF162 sequences. Figure 66B presents alignments of modified US4 sequences. The SEQ ID NOS for these sequences are presented in Tables 1A and 1B.

Figure 67 shows the ELISA titers (binding antibodies) obtained in two rhesus macaques (H445, lines with solid black dots; and J408, lines with open squares). The y-axis is the end-point gp140 ELISA titers and the x-axis shows weeks post-immunization. The dashed lines at 0, 4, and 8 weeks represent DNA immunizations. The alternating dash/dotted line at 27 weeks indicates a DNA plus protein boost immunization.

Figure 68 (SEQ ID NO:77) depicts the wild-type nucleotide sequence of Gag reverse transcriptase from SF2.

Figure 69 (SEQ ID NO:78) depicts the nucleotide sequence of the construct designated GP1.

Figure 70 (SEQ ID NO:79) depicts the nucleotide sequence of the construct designated GP2.

Figure 71 (SEQ ID NO:80) depicts the nucleotide sequence of the construct designated FS(+).protinact.RTopt.YM. FS(+) indicates that there is a frameshift in the GagPol coding sequence.

Figure 72 (SEQ ID NO:81) depicts the nucleotide sequence of the construct designated FS(+).protinact.RTopt.YMWM.

Figure 73 (SEQ ID NO:82) depicts the nucleotide sequence of the construct designated FS(-

).protmod.RTopt.YM. FS(-) indicates that there is no frameshift in the GagPol coding sequence.

Figure 74 (SEQ ID NO:83) depicts the nucleotide sequence of the construct designated

5 FS(-).protmod.RTopt.YMWM.

Figure 75 (SEQ ID NO:84) depicts the nucleotide sequence of the construct designated FS(-).protmod.RTopt(+).

Figure 76 (SEQ ID NO:85) depicts the nucleotide sequence of wild type Tat from isolate SF162.

10 Figure 77 (SEQ ID NO:86) depicts the amino acid sequence of the tat polypeptide.

Figure 78 (SEQ ID NO:87) depicts the nucleotide sequence of a synthetic Tat construct designated

15 Tat.SF162.opt.

Figure 79 (SEQ ID NO:88) depicts the nucleotide sequence of a synthetic Tat construct designated tat.cys22.sf162.opt. The construct encodes a tat polypeptide in which the cystein residue at position 22

20 of the wild type Tat polypeptide is replaced by a glycine residue.

Figures 80A to 80E are an alignment of the nucleotide sequences of the constructs designated Gag.mod.SF2, GP1 (SEQ ID NO:78), and GP2 (SEQ ID NO:79).

25 Figure 81 (SEQ ID NO:89) depicts the nucleotide sequence of the construct designated tataminoSF162.opt, which encodes the amino terminus of that tat protein. The codon encoding the cystein-22 residue is underlined.

30 Figure 82 (SEQ ID NO:90) depicts the amino acid sequence of the polypeptide encoded by the construct designated tat.cys22.SF162.opt (SEQ ID NO:88).

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag).

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

25

1. DEFINITIONS

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

"Synthetic" sequences, as used herein, refers to Env-, tat- or Gag-encoding polynucleotides whose expression has been optimized as described herein, for example, by codon substitution, deletions, replacements and/or inactivation of inhibitory sequences. "Wild-type"

or "native" sequences, as used herein, refers to polypeptide encoding sequences that are essentially as they are found in nature, e.g., Gag encoding sequences as found in the isolate HIV-1SF2 or Env encoding sequences as found in the isolates HIV-1SF162 or HIV1US4.

As used herein, the term "virus-like particle" or "VLP" refers to a nonreplicating, viral shell, derived from any of several viruses discussed further below. VLPs are generally composed of one or more viral proteins, such as, but not limited to those proteins referred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming polypeptides derived from these proteins. VLPs can form spontaneously upon recombinant expression of the protein in an appropriate expression system. Methods for producing particular VLPs are known in the art and discussed more fully below. The presence of VLPs following recombinant expression of viral proteins can be detected using conventional techniques known in the art, such as by electron microscopy, biophysical characterization, and the like. See, e.g., Baker et al., *Biophys. J.* (1991) 60:1445-1456; Hagensee et al., *J. Virol.* (1994) 68:4503-4505. For example, VLPs can be isolated by density gradient centrifugation and/or identified by characteristic density banding (e.g., Example 7). Alternatively, cryoelectron microscopy can be performed on vitrified aqueous samples of the VLP preparation in question, and images recorded under appropriate exposure conditions.

By "particle-forming polypeptide" derived from a particular viral protein is meant a full-length or near full-length viral protein, as well as a fragment thereof, or a viral protein with internal deletions, which has the ability to form VLPs under conditions that favor VLP

formation. Accordingly, the polypeptide may comprise the full-length sequence, fragments, truncated and partial sequences, as well as analogs and precursor forms of the reference molecule. The term therefore intends 5 deletions, additions and substitutions to the sequence, so long as the polypeptide retains the ability to form a VLP. Thus, the term includes natural variations of the specified polypeptide since variations in coat proteins often occur between viral isolates. The term also 10 includes deletions, additions and substitutions that do not naturally occur in the reference protein, so long as the protein retains the ability to form a VLP. Preferred substitutions are those which are conservative in nature, i.e., those substitutions that take place within a family 15 of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline, 20 phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cystine, serine threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids.

25 An "antigen" refers to a molecule containing one or more epitopes (either linear, conformational or both) that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is used interchangeably with the term "immunogen." 30 Normally, a B-cell epitope will include at least about 5 amino acids but can be as small as 3-4 amino acids. A T-cell epitope, such as a CTL epitope, will include at least about 7-9 amino acids, and a helper T-cell epitope at least about 12-20 amino acids. Normally, an epitope

will include between about 7 and 15 amino acids, such as, 9, 10, 12 or 15 amino acids. The term "antigen" denotes both subunit antigens, (i.e., antigens which are separate and discrete from a whole organism with which the antigen 5 is associated in nature), as well as, killed, attenuated or inactivated bacteria, viruses, fungi, parasites or other microbes. Antibodies such as anti-idiotype antibodies, or fragments thereof, and synthetic peptide mimotopes, which can mimic an antigen or antigenic 10 determinant, are also captured under the definition of antigen as used herein. Similarly, an oligonucleotide or polynucleotide which expresses an antigen or antigenic determinant *in vivo*, such as in gene therapy and DNA immunization applications, is also included in the 15 definition of antigen herein.

For purposes of the present invention, antigens can be derived from any of several known viruses, bacteria, parasites and fungi, as described more fully below. The term also intends any of the various tumor antigens. 20 Furthermore, for purposes of the present invention, an "antigen" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the 25 ability to elicit an immunological response, as defined herein. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the antigens.

30 An "immunological response" to an antigen or composition is the development in a subject of a humoral and/or a cellular immune response to an antigen present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to

an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTL's"). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the

antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376. Recent methods of 5 measuring cell-mediated immune response include measurement of intracellular cytokines or cytokine secretion by T-cell populations, or by measurement of epitope specific T-cells (e.g., by the tetramer technique) (reviewed by McMichael, A.J., and O'Callaghan, 10 C.A., *J. Exp. Med.* 187(9):1367-1371, 1998; McHeyzer-Williams, M.G., et al., *Immunol. Rev.* 150:5-21, 1996; Lalvani, A., et al., *J. Exp. Med.* 186:859-865, 1997).

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or 15 the production or activation of helper T-cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the 20 activation of suppressor T-cells and/or $\gamma\delta$ T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate antibody-complement, or antibody dependent cell 25 cytotoxicity (ADCC) to provide protection to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

An "immunogenic composition" is a composition that 30 comprises an antigenic molecule where administration of the composition to a subject results in the development in the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest.

By "subunit vaccine" is meant a vaccine composition which includes one or more selected antigens but not all antigens, derived from or homologous to, an antigen from a pathogen of interest such as from a virus, bacterium, parasite or fungus. Such a composition is substantially free of intact pathogen cells or pathogenic particles, or the lysate of such cells or particles. Thus, a "subunit vaccine" can be prepared from at least partially purified (preferably substantially purified) immunogenic polypeptides from the pathogen, or analogs thereof. The method of obtaining an antigen included in the subunit vaccine can thus include standard purification techniques, recombinant production, or synthetic production.

"Substantially purified" generally refers to isolation of a substance (compound, polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample a substantially purified component comprises 50%, preferably 80%-85%, more preferably 90-95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and sedimentation according to density.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but

is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral or procaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to
5 the coding sequence.

Typical "control elements", include, but are not limited to, transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences (located 3' to the translation
10 stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), and translation termination sequences, see e.g., McCaughan et al. (1995) *PNAS USA* 92:5431-5435; Kochetov et al (1998) *FEBS Letts.* 440:351-355.

15 A "nucleic acid" molecule can include, but is not limited to, procaryotic sequences, eucaryotic mRNA, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. The term also captures sequences that include
20 any of the known base analogs of DNA and RNA.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter operably linked to a coding sequence is
25 capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet
30 transcribed sequences can be present between the promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting prokaryotic microorganisms or eucaryotic cell lines cultured as unicellular entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

Techniques for determining amino acid sequence "similarity" are well known in the art. In general, "similarity" means the exact amino acid to amino acid comparison of two or more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. A so-termed "percent similarity" then

can be determined between the compared polypeptide sequences. Techniques for determining nucleic acid and amino acid sequence identity also are well known in the art and include determining the nucleotide sequence of 5 the mRNA for that gene (usually via a cDNA intermediate) and determining the amino acid sequence encoded thereby, and comparing this to a second amino acid sequence. In general, "identity" refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of 10 two polynucleotides or polypeptide sequences, respectively.

Two or more polynucleotide sequences can be compared by determining their "percent identity." Two or more amino acid sequences likewise can be compared by 15 determining their "percent identity." The percent identity of two sequences, whether nucleic acid or peptide sequences, is generally described as the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. 20 An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be extended to use with peptide sequences using the scoring matrix developed by 25 Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An implementation of this algorithm for nucleic 30 acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in their BestFit utility application. The default parameters for this method are

described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). Other equally suitable programs for calculating the percent identity or
5 similarity between sequences are generally known in the art.

For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions. Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh,
10 developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages, the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated, the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, such as the alignment program
15 BLAST, which can also be used with default parameters. For example, BLASTN and BLASTP can be used with the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by =
20 HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at
25

the following internet address:

<http://www.ncbi.nlm.gov/cgi-bin/BLAST>.

One of skill in the art can readily determine the proper search parameters to use for a given sequence in 5 the above programs. For example, the search parameters may vary based on the size of the sequence in question. Thus, for example, a representative embodiment of the present invention would include an isolated polynucleotide having X contiguous nucleotides, wherein 10 (i) the X contiguous nucleotides have at least about 50% identity to Y contiguous nucleotides derived from any of the sequences described herein, (ii) X equals Y, and (iii) X is greater than or equal to 6 nucleotides and up to 5000 nucleotides, preferably greater than or equal to 15 15 8 nucleotides and up to 5000 nucleotides, more preferably 10-12 nucleotides and up to 5000 nucleotides, and even more preferably 15-20 nucleotides, up to the number of nucleotides present in the full-length sequences described herein (e.g., see the Sequence Listing and 20 claims), including all integer values falling within the above-described ranges.

The synthetic expression cassettes (and purified polynucleotides) of the present invention include related polynucleotide sequences having about 80% to 100%, 25 greater than 80-85%, preferably greater than 90-92%, more preferably greater than 95%, and most preferably greater than 98% sequence (including all integer values falling within these described ranges) identity to the synthetic expression cassette sequences disclosed herein (for 30 example, to the sequences presented in Tables 1A and 1B) when the sequences of the present invention are used as the query sequence.

Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" typically

hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected 5 nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe.

10 Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames 15 and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a 20 particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or 25 absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions 30 is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A

Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

A first polynucleotide is "derived from" second polynucleotide if it has the same or substantially the 5 same basepair sequence as a region of the second polynucleotide, its cDNA, complements thereof, or if it displays sequence identity as described above.

A first polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first 10 polynucleotide derived from a second polynucleotide, or (ii) displays sequence identity to the second polypeptides as described above.

Generally, a viral polypeptide is "derived from" a particular polypeptide of a virus (viral polypeptide) if 15 it is (i) encoded by an open reading frame of a polynucleotide of that virus (viral polynucleotide), or (ii) displays sequence identity to polypeptides of that virus as described above.

"Encoded by" refers to a nucleic acid sequence which 20 codes for a polypeptide sequence, wherein the polypeptide sequence or a portion thereof contains an amino acid sequence of at least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even more preferably at least 15 to 20 amino acids from a polypeptide encoded by 25 the nucleic acid sequence. Also encompassed are polypeptide sequences which are immunologically identifiable with a polypeptide encoded by the sequence.

"Purified polynucleotide" refers to a polynucleotide of interest or fragment thereof which is essentially 30 free, e.g., contains less than about 50%, preferably less than about 70%, and more preferably less than about 90%, of the protein with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides of interest are well-known in the art and include, for

example, disruption of the cell containing the polynucleotide with a chaotropic agent and separation of the polynucleotide(s) and proteins by ion-exchange chromatography, affinity chromatography and sedimentation according to density.

By "nucleic acid immunization" is meant the introduction of a nucleic acid molecule encoding one or more selected antigens into a host cell, for the *in vivo* expression of an antigen, antigens, an epitope, or epitopes. The nucleic acid molecule can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have been removed from the host. In the latter case, the transformed cells are reintroduced into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

"Gene transfer" or "gene delivery" refers to methods or systems for reliably inserting DNA or RNA of interest into a host cell. Such methods can result in transient expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred replicons (e.g., episomes), or integration of transferred genetic material into the genomic DNA of host cells. Gene delivery expression vectors include, but are not limited to, vectors derived from bacterial plasmid vectors, viral vectors, non-viral vectors, alphaviruses, pox viruses and vaccinia viruses. When used for immunization, such gene delivery expression vectors may be referred to as vaccines or vaccine vectors.

"T lymphocytes" or "T cells" are non-antibody producing lymphocytes that constitute a part of the cell-mediated arm of the immune system. T cells arise from immature lymphocytes that migrate from the bone marrow to

the thymus, where they undergo a maturation process under the direction of thymic hormones. Here, the mature lymphocytes rapidly divide increasing to very large numbers. The maturing T cells become immunocompetent based on their ability to recognize and bind a specific antigen. Activation of immunocompetent T cells is triggered when an antigen binds to the lymphocyte's surface receptors.

The term "transfection" is used to refer to the uptake of foreign DNA by a cell. A cell has been "transfected" when exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are generally known in the art. See, e.g., Graham et al. (1973) *Virology*, 52:456, Sambrook et al. (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) *Basic Methods in Molecular Biology*, Elsevier, and Chu et al. (1981) *Gene* 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells. The term refers to both stable and transient uptake of the genetic material, and includes uptake of peptide- or antibody-linked DNAs.

A "vector" is capable of transferring gene sequences to target cells (e.g., bacterial plasmid vectors, viral vectors, non-viral vectors, particulate carriers, and liposomes). Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

Transfer of a "suicide gene" (e.g., a drug-susceptibility gene) to a target cell renders the cell sensitive to compounds or compositions that are

relatively nontoxic to normal cells. Moolten, F.L. (1994) *Cancer Gene Ther.* 1:279-287. Examples of suicide genes are thymidine kinase of herpes simplex virus (HSV-tk), cytochrome P450 (Manome et al. (1996) *Gene Therapy* 3:513-520), human deoxycytidine kinase (Manome et al. 5 (1996) *Nature Medicine* 2(5):567-573) and the bacterial enzyme cytosine deaminase (Dong et al. (1996) *Human Gene Therapy* 7:713-720). Cells which express these genes are rendered sensitive to the effects of the relatively 10 nontoxic prodrugs ganciclovir (HSV-tk), cyclophosphamide (cytochrome P450 2B1), cytosine arabinoside (human deoxycytidine kinase) or 5-fluorocytosine (bacterial cytosine deaminase). Culver et al. (1992) *Science* 256:1550-1552, Huber et al. (1994) *Proc. Natl. Acad. Sci.* 15 USA 91:8302-8306.

A "selectable marker" or "reporter marker" refers to a nucleotide sequence included in a gene transfer vector that has no therapeutic activity, but rather is included to allow for simpler preparation, manufacturing, 20 characterization or testing of the gene transfer vector.

A "specific binding agent" refers to a member of a specific binding pair of molecules wherein one of the molecules specifically binds to the second molecule through chemical and/or physical means. One example of a 25 specific binding agent is an antibody directed against a selected antigen.

By "subject" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such 30

as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The system described above 5 is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

By "pharmaceutically acceptable" or "pharmacologically acceptable" is meant a material which 10 is not biologically or otherwise undesirable, i.e., the material may be administered to an individual in a formulation or composition without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the 15 composition in which it is contained.

By "physiological pH" or a "pH in the physiological range" is meant a pH in the range of approximately 7.2 to 8.0 inclusive, more typically in the range of approximately 7.2 to 7.6 inclusive.

20 As used herein, "treatment" refers to any of (I) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may 25 be effected prophylactically (prior to infection) or therapeutically (following infection).

"Lentiviral vector", and "recombinant lentiviral vector" are derived from the subset of retroviral vectors known as lentiviruses. Lentiviral vectors refer to a 30 nucleic acid construct which carries, and within certain embodiments, is capable of directing the expression of a nucleic acid molecule of interest. The lentiviral vector includes at least one transcriptional promoter/enhancer or locus defining element(s), or other elements which

control gene expression by other means such as alternate splicing, nuclear RNA export, post-translational modification of messenger, or post-transcriptional modification of protein. Such vector constructs must

5 also include a packaging signal, long terminal repeats (LTRS) or portion thereof, and positive and negative strand primer binding sites appropriate to the lentiviral vector used (if these are not already present in the retroviral vector). Optionally, the recombinant

10 lentiviral vector may also include a signal which directs polyadenylation, selectable markers such as Neo, TK, hygromycin, phleomycin, histidinol, or DHFR, as well as one or more restriction sites and a translation termination sequence. By way of example, such vectors

15 typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second strand DNA synthesis, and a 3'LTR or a portion thereof.

"Lentiviral vector particle" as utilized within the present invention refers to a lentivirus which carries at least one gene of interest. The retrovirus may also contain a selectable marker. The recombinant lentivirus is capable of reverse transcribing its genetic material (RNA) into DNA and incorporating this genetic material into a host cell's DNA upon infection. Lentiviral vector particles may have a lentiviral envelope, a non-lentiviral envelope (e.g., an amphi or VSV-G envelope), or a chimeric envelope.

"Nucleic acid expression vector" or "Expression cassette" refers to an assembly which is capable of directing the expression of a sequence or gene of interest. The nucleic acid expression vector includes a promoter which is operably linked to the sequences or gene(s) of interest. Other control elements may be present as well. Expression cassettes described herein

may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include a bacterial origin of replication, one or more selectable markers, a signal 5 which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 origin of replication), a multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

10 "Packaging cell" refers to a cell which contains those elements necessary for production of infectious recombinant retrovirus (e.g., lentivirus) which are lacking in a recombinant retroviral vector. Typically, such packaging cells contain one or more expression 15 cassettes which are capable of expressing proteins which encode Gag, pol and env proteins.

"Producer cell" or "vector producing cell" refers to a cell which contains all elements necessary for production of recombinant retroviral vector particles.

20

2. MODES OF CARRYING OUT THE INVENTION

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such 25 may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of methods and materials similar 30 or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

2.1. SYNTHETIC EXPRESSION CASSETTES**2.1.1 MODIFICATION OF HIV-1 GAG NUCLEIC ACID CODING SEQUENCES**

One aspect of the present invention is the generation of HIV-1 Gag protein coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequence. An exemplary embodiment of the present invention is illustrated herein modifying the Gag protein wild-type sequences obtained from the HIV-1SF2 strain (SEQ ID NO:1; Sanchez-Pescador, R., et al., *Science* 227(4686): 484-492, 1985; Luciw, P.A., et al. U.S. Patent No. 5,156,949, issued October 20, 1992; Luciw, P.A., et al., U.S. Patent No. 5,688,688, November 18, 1997). Gag sequence obtained from other HIV variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Gag protein encoding sequences obtained from the isolates HIV_{IIIB}, HIV_{SF2}, HIV-
1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{uc1} and HIV-2_{uc2}), and simian immunodeficiency virus (SIV). (See, e.g., *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA; for a description of these and other related viruses).

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet.

The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the 5 nucleotides G or C. The Gag coding sequences were modified to be comparable to codon usage found in highly expressed human genes. In Figure 11 (Example 1), the percent A-T content of cDNA sequences corresponding to the mRNA for a known unstable mRNA and a known stable 10 mRNA are compared to the percent A-T content of native HIV-1SF2 Gag cDNA and to the synthetic Gag cDNA sequence of the present invention. Experiments performed in support of the present invention showed that the synthetic Gag sequences were capable of higher level of 15 protein production (see the Examples) relative to the native Gag sequences. The data in Figure 11 suggest that one reason for this increased production is increased stability of the mRNA corresponding to the synthetic Gag coding sequences versus the mRNA corresponding to the 20 native Gag coding sequences.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag coding sequences (Example 1). The RRE is a secondary RNA structure that interacts with the HIV encoded Rev- 25 protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements were inactivated by introducing multiple point mutations that did not alter the reading frame of the 30 encoded proteins. Figure 1 shows the original SF2 Gag sequence, the location of the INS sequences, and the modifications made to the INS sequences to reduce their effects. The resulting modified coding sequences are

presented as a synthetic Gag expression cassette (SEQ ID NO:4).

Modification of the Gag polypeptide coding sequences resulted in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Further, expression of the sequences resulted in production of virus-like particles (VLPs) by these cell lines (see below). Similar Gag polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, strains, etc.) including, but not limited to such other variants include, but are not limited to, Gag polypeptide encoding sequences obtained from the isolates HIV_{IIIB}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}), and simian immunodeficiency virus (SIV). (See, e.g., Virology, 3rd Edition (W.K. Joklik ed. 1988); Fundamental Virology, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991; Virology, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA). Gag polypeptide encoding sequences derived from these variants can be optimized and tested for improved expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 1).

2.1.2 FURTHER MODIFICATION OF SEQUENCES INCLUDING HIV-1 GAG NUCLEIC ACID CODING SEQUENCES

Experiments performed in support of the present invention have shown that similar modifications of HIV-1 Gag-protease, Gag-reverse transcriptase and Gag-polymerase sequences also result in improved expression

of the polyproteins, as well as, the production of VLPs formed by polypeptides produced from such modified coding sequences.

For the Gag-protease sequence (wild type, SEQ ID NO:2; modified, SEQ ID NOS:5, 78, 79), the changes in codon usage were restricted to the regions upstream of the -1 frameshift (Figure 2). Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence were altered as well (indicated in Figure 2). Exemplary constructs (which include the -1 frameshift) encoding modified Gag-protease sequences include those shown in SEQ ID NOS:78 and 79 (Figures 69 and 70). These are: GP1 (SEQ ID NO:78) in which the protease region was also codon optimized and INS inactivated and GP2 (SEQ ID NO:79), in which the protease region was only subjected to INS inactivation.

For other Gag-containing sequences, for example the Gag-polymerase sequence (wild type, SEQ ID NO:3; modified, SEQ ID NO:6) or Gag-reverse transcriptase (wild type, SEQ ID NO:77; modified SEQ ID NOS:80-84), the changes in codon usage are similar to those for the Gag-protease sequence. Those expression cassettes which contain a frameshift in the GagPol coding sequence are designated "FS(+)" (SEQ ID NOS:80 and 81, Figures 71 and 72) while the designation "FS(-)" (SEQ ID Nos: 82, 83 and 84, Figures 73, 74 and 75) indicates that there is no frameshift utilized in this coding sequence.

In addition to polyproteins containing HIV-related sequences, the various Gag-, Gag-prot, Gag-pol, Gag-reverse transcriptase encoding sequences of the present invention can be fused to other polypeptides (creating chimeric polypeptides) for which an immunogenic response is desired. An example of such a chimeric protein is the

joining of the improved expression Gag encoding sequences to the Hepatitis C Virus (HCV) core protein. In this case, the HCV-core encoding sequences were placed in-frame with the HIV-Gag encoding sequences, resulting in 5 the Gag/HCV-core encoding sequence presented as SEQ ID NO:7 (wild type sequence presented as SEQ ID NO:8).

Further sequences useful in the practice of the present invention include, but are not limited to, sequences encoding viral epitopes/antigens {including but 10 not limited to, HCV antigens (e.g., E1, E2; Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; 15 Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997), HIV antigens (e.g., derived from nef, tat, rev, vpu, vif, 20 vpr and/or env); and sequences encoding tumor antigens/epitopes. Additional sequences are described below. Also, variations on the orientation of the Gag and other coding sequences, relative to each other, are also described below.

25 Gag, Gag-protease, Gag-reverse transcriptase and/or Gag-polymerase polypeptide coding sequences can be obtained from any HIV isolates (different families, subtypes, and strains) including but not limited to the isolates HIV_{IIIB}, HIV_{SF2}, HIV_{SF162}, HIV_{us4}, HIV_{cm235}, HIV_{LAV}, 30 HIV_{LAI}, HIV_{MN}) (see, e.g., Myers et al. Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., Human Retroviruses and Aids, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory). Synthetic expression cassettes can be generated using

such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1). Further, the synthetic expression cassettes of the present invention include related Gag polypeptide coding sequences having greater than 75%, preferably greater than 80-85%, more preferably greater than 90-95%, and most preferably greater than 98% sequence identity (or any integer value within these ranges) to the synthetic expression cassette sequences disclosed herein (for example, SEQ ID NO:4; SEQ ID NO:5; SEQ ID NO:6; and SEQ ID NO:20, the Gag Major Homology Region).

2.1.3 EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1

GAG AND RELATED POLYPEPTIDES

Several synthetic Gag-encoding sequences (expression cassettes) of the present invention were cloned into a number of different expression vectors (Example 1) to evaluate levels of expression and production of VLPs. Two modified synthetic coding sequences are presented as

a synthetic Gag expression cassette (SEQ ID NO:4) and a synthetic Gag-protease expression cassette (SEQ ID NOS:78 and 79). Other synthetic Gag-encoding proteins are presented, for example, as SEQ ID NOS:80 through 84. The synthetic DNA fragments for Gag-encoding polypeptides (e.g., Gag, Gag-protease, Gag-polymerase, Gag-reverse transcriptase) were cloned into expression vectors described in Example 1, including, a transient expression vector, CMV-promoter-based mammalian vectors, and a shuttle vector for use in baculovirus expression systems. Corresponding wild-type sequences were cloned into the same vectors.

These vectors were then transfected into a several different cell types, including a variety of mammalian

cell lines, (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines were cultured under appropriate conditions and the levels of p24 (Gag) expression in supernatants were evaluated (Example 2). The results of these assays demonstrated that expression of synthetic Gag-encoding sequences were significantly higher than corresponding wild-type sequences (Example 2; Table 2).

Further, Western Blot analysis showed that cells containing the synthetic Gag expression cassette produced the expected 55 kD (p55) protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag p55 protein was seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants for cells transfected with the synthetic Gag expression cassette of the present invention. Experiments performed in support of the present invention suggest that cells containing the synthetic Gag-prot expression cassettes produced the expected Gag-prot protein at comparably higher per-cell concentrations than cells containing the wild-type expression cassette.

Fractionation of the supernatants from mammalian cells transfected with the synthetic Gag expression cassette showed that it provides superior production of both p55 protein and VLPs, relative to the wild-type Gag sequences (Examples 6 and 7).

Efficient expression of these Gag-containing polypeptides in mammalian cell lines provides the following benefits: the Gag polypeptides are free of baculovirus contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Gag-containing polypeptides in

CHO or other mammalian cells which is not feasible in the absence of the increased expression obtained using the constructs of the present invention. Exemplary Mammalian cell lines include, but are not limited to, BHK, VERO, 5 HT1080, 293, 293T, RD, COS-7, CHO, Jurkat, HUT, SUPT, C8166, MOLT4/clone8, MT-2, MT-4, H9, PM1, CEM, myeloma cells (e.g., SB20 cells) and CEMX174, such cell lines are available, for example, from the A.T.C.C.).

A synthetic Gag expression cassette of the present 10 invention also demonstrated high levels of expression and VLP production when transfected into insect cells (Example 7). Further, in addition to a higher total protein yield, the final product from the synthetic p55-expressed Gag consistently contained lower amounts of 15 contaminating baculovirus proteins than the final purified product from the native p55-expressed Gag.

Further, synthetic Gag expression cassettes of the present invention have also been introduced into yeast vectors which were transformed into and efficiently 20 expressed by yeast cells (*Saccharomyces cerevisiae*; using vectors as described in Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998).

In addition to the mammalian and insect vectors 25 described in the Examples, the synthetic expression cassettes of the present invention can be incorporated into a variety of expression vectors using selected expression control elements. Appropriate vectors and control elements for any given cell type can be selected by one having ordinary skill in the art in view of the 30 teachings of the present specification and information known in the art about expression vectors.

For example, a synthetic Gag expression cassette can be inserted into a vector which includes control elements operably linked to the desired coding sequence, which

allow for the expression of the gene in a selected cell-type. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter (a CMV promoter can include intron A), RSV, HIV-LTR, the mouse mammary tumor virus LTR promoter (MMLV-LTR), FIV-LTR, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook, et al., *supra*, as well as a bovine growth hormone terminator sequence. Introns, containing splice donor and acceptor sites, may also be designed into the constructs for use with the present invention (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs.

Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

The desired synthetic Gag polypeptide encoding sequences can be cloned into any number of commercially available vectors to generate expression of the polypeptide in an appropriate host system. These systems include, but are not limited to, the following:

5 baculovirus expression {Reilly, P.R., et al., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992); Beames, et al., Biotechniques 11:378 (1991); Pharmingen; Clontech, Palo Alto, CA}), vaccinia expression {Earl, P. L., et al.,

10 "Expression of proteins in mammalian cells using vaccinia" In *Current Protocols in Molecular Biology* (F. M. Ausubel, et al. Eds.), Greene Publishing Associates & Wiley Interscience, New York (1991); Moss, B., et al., U.S. Patent Number 5,135,855, issued 4 August 1992},

15 expression in bacteria {Ausubel, F.M., et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA; Clontech}, expression in yeast {Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998; Shuster, J.R., U.S. Patent No. 5,629,203,

20 issued May 13, 1997; Gellissen, G., et al., Antonie Van Leeuwenhoek, 62(1-2):79-93 (1992); Romanos, M.A., et al., Yeast 8(6):423-488 (1992); Goeddel, D.V., *Methods in Enzymology* 185 (1990); Guthrie, C., and G.R. Fink, *Methods in Enzymology* 194 (1991)}; expression in

25 mammalian cells {Clontech; Gibco-BRL, Ground Island, NY; e.g., Chinese hamster ovary (CHO) cell lines (Haynes, J., et al., Nuc. Acid. Res. 11:687-706 (1983); 1983, Lau, Y.F., et al., Mol. Cell. Biol. 4:1469-1475 (1984); Kaufman, R. J., "Selection and coamplification of

30 heterologous genes in mammalian cells," in *Methods in Enzymology*, vol. 185, pp537-566. Academic Press, Inc., San Diego CA (1991)}, and expression in plant cells {plant cloning vectors, Clontech Laboratories, Inc., Palo Alto, CA, and Pharmacia LKB Biotechnology, Inc.,

Piscataway, NJ; Hood, E., et al., *J. Bacteriol.* 168:1291-1301 (1986); Nagel, R., et al., *FEMS Microbiol. Lett.* 67:325 (1990); An, et al., "Binary Vectors", and others in Plant Molecular Biology Manual A3:1-19 (1988);
5 Miki, B.L.A., et al., pp.249-265, and others in Plant DNA Infectious Agents (Hohn, T., et al., eds.) Springer-Verlag, Wien, Austria, (1987); *Plant Molecular Biology: Essential Techniques*, P.G. Jones and J.M. Sutton, New York, J. Wiley, 1997; Miglani, Gurbachan *Dictionary of Plant Genetics and Molecular Biology*, New York, Food Products Press, 1998; Henry, R. J., *Practical Applications of Plant Molecular Biology*, New York, Chapman & Hall, 1997}.

Also included in the invention is an expression vector, such as the CMV promoter-containing vectors described in Example 1, containing coding sequences and expression control elements which allow expression of the coding regions in a suitable host. The control elements generally include a promoter, translation initiation codon, and translation and transcription termination sequences, and an insertion site for introducing the insert into the vector. Translational control elements have been reviewed by M. Kozak (e.g., Kozak, M., *Mamm. Genome* 7(8):563-574, 1996; Kozak, M., *Biochimie* 76(9):815-821, 1994; Kozak, M., *J Cell Biol* 108(2):229-241, 1989; Kozak, M., and Shatkin, A.J., *Methods Enzymol* 60:360-375, 1979).

Expression in yeast systems has the advantage of commercial production. Recombinant protein production by 30 vaccinia and CHO cell line have the advantage of being mammalian expression systems. Further, vaccinia virus expression has several advantages including the following: (i) its wide host range; (ii) faithful post-

WO 00/39302

transcriptional modification, processing, folding, transport, secretion, and assembly of recombinant proteins; (iii) high level expression of relatively soluble recombinant proteins; and (iv) a large capacity to accommodate foreign DNA.

The recombinantly expressed polypeptides from synthetic Gag-encoding expression cassettes are typically isolated from lysed cells or culture media. Purification can be carried out by methods known in the art including salt fractionation, ion exchange chromatography, gel filtration, size-exclusion chromatography, size-fractionation, and affinity chromatography.

Immunoaffinity chromatography can be employed using antibodies generated based on, for example, Gag antigens.

Advantages of expressing the Gag-containing proteins of the present invention using mammalian cells include, but are not limited to, the following: well-established protocols for scale-up production; the ability to produce VLPs; cell lines are suitable to meet good manufacturing process (GMP) standards; culture conditions for mammalian cells are known in the art.

2.1.4 MODIFICATION OF HIV-1 ENV NUCLEIC ACID CODING

SEQUENCES

One aspect of the present invention is the generation of HIV-1 Env protein coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequence. Exemplary embodiments of the present invention are illustrated herein modifying the Env protein wild-type sequences obtained from the HIV-1 subtype B strains HIV-1US4 and HIV-1SF162 (Myers et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., Human Retroviruses and Aids, 1997, Los Alamos,

New Mexico: Los Alamos National Laboratory). Env sequence obtained from other HIV variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include those
5 described above in Section 2.1.1 and on the World Wide Web (Internet), for example at http://hiv-web.lanl.gov/cgi-bin/hivDB3/public/wdb/ssampublic and http://hiv-web.lanl.gov.

First, the HIV-1 codon usage pattern was modified so
10 that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content
15 in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Env coding sequences were modified to be comparable to codon usage found in highly
20 expressed human genes. Experiments performed in support of the present invention showed that the synthetic Env sequences were capable of higher level of protein production (see the Examples) relative to the native Env sequences. One reason for this increased production may
25 be increased stability of the mRNA corresponding to the synthetic Env coding sequences versus the mRNA corresponding to the native Env coding sequences.

Modification of the Env polypeptide coding sequences resulted in improved expression relative to the wild-type
30 coding sequences in a number of mammalian cell lines. Similar Env polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, etc.). Env polypeptide encoding sequences derived from these variants can be optimized and tested for improved

expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 2).

5 2.1.5 **FURTHER MODIFICATION OF HIV-1 ENV NUCLEAR ACID
CODING SEQUENCES**

In addition to proteins containing HIV-related sequences, the Env encoding sequences of the present invention can be fused to other polypeptides (creating 10 chimeric polypeptides). Also, variations on the orientation of the Env and other coding sequences, relative to each other, are contemplated. Further, the HIV protein encoding cassettes of the present invention can be co-expressed using one vector or multiple vectors. 15 In addition, the polyproteins can be operably linked to the same or different promoters.

Env polypeptide coding sequences can be obtained from any HIV isolates (different families, subtypes, and strains) including but not limited to the isolates HIV_{IIIB}, 20 HIV_{SF2}, HIV_{us4}, HIV_{CM235}, HIV_{SF162}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}) (see, e.g., Myers et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory). Synthetic 25 expression cassettes can be generated using such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1). Further, the synthetic expression cassettes (and purified polynucleotides) of the present invention include related 30 Env polypeptide coding sequences having greater than 90%, preferably greater than 92%, more preferably greater than 95%, and most preferably greater than 98% sequence identity to the synthetic expression cassette sequences disclosed herein (for example, SEQ ID NOS:71-72; and/or

the sequences presented in Tables 1A and 1B) when the sequences of the present invention are used as the query sequence.

5 2.1.6 **EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1
ENV AND RELATED POLYPEPTIDES**

Several synthetic Env-encoding sequences (expression cassettes) of the present invention were cloned into a number of different expression vectors (Example 1) to 10 evaluate levels of expression and production of Env polypeptide. A modified synthetic coding sequence is presented as synthetic Env expression cassettes (Example 1, e.g., Tables 1A and 1B). The synthetic DNA fragments for Env were cloned into eucaryotic expression vectors 15 described in Example 1 and in Section 2.1.3 above, including, a transient expression vector and CMV-promoter-based mammalian vectors. Corresponding wild-type sequences were cloned into the same vectors.

These vectors were then transfected into a several 20 different cell types, including a variety of mammalian cell lines, (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines were cultured under appropriate conditions and the levels of gp120, gp140 and gp160 Env expression in 25 supernatants were evaluated (Example 2). Env polypeptides include, but are not limited to, for example, native gp160, oligomeric gp140, monomeric gp120 as well as modified sequences of these polypeptides. The results of these assays demonstrated that expression of 30 synthetic Env encoding sequences were significantly higher than corresponding wild-type sequences (Example 2; Tables 3 and 4).

Further, Western Blot analysis showed that cells containing the synthetic Env expression cassette produced

WO 00/39302

the expected protein (gp120, gp140 or gp160) at higher per-cell concentrations than cells containing the native expression cassette. The Env proteins were seen in both cell lysates and supernatants. The levels of production 5 were significantly higher in cell supernatants for cells transfected with the synthetic Env expression cassettes of the present invention as compared to wild type.

Fractionation of the supernatants from mammalian 10 cells transfected with the synthetic Env expression cassettes showed that it provides superior production of Env proteins, relative to the wild-type Env sequences (Examples 2 and 3).

Efficient expression of these Env-containing 15 polypeptides in mammalian cell lines provides the following benefits: the Env polypeptides are free of baculovirus or other viral contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Env- 20 containing polypeptides in CHO cells which is less feasible in the absence of the increased expression obtained using the constructs of the present invention.

Exemplary cell lines (e.g., mammalian, yeast, insect, etc.) include those described above in Section 25 2.1.3 for Gag-containing constructs. Further, appropriate vectors and control elements (e.g., promoters, enhancers, polyadenylation sequences, etc.) for any given cell type can be selected, as described above in Section 2.1.3, by one having ordinary skill in the art in view of the 30 teachings of the present specification and information known in the art about expression vectors. In addition, the recombinantly expressed polypeptides from synthetic Env-encoding expression cassettes are typically isolated and purified from lysed cells or culture media, as

described above for Gag-encoding expression cassettes. An exemplary purification is described in Example 4 and shown in Figure 60.

5 2.1.7 MODIFICATION OF HIV-1 TAT NUCLEAR ACID CODING
SEQUENCES

Another aspect of the present invention is the generation of HIV-1 tat protein coding sequences, and related sequences, having improved expression relative to 10 the corresponding wild-type sequence. Exemplary embodiments of the present invention are illustrated herein modifying the tat wild-type nucleotide sequence (SEQ ID NO:85, Figure 76) obtained from SF162 as described above. Exemplary synthetic tat constructs are 15 shown in SEQ ID NO:87, which depicts a tat construct encoding a full-length tat polypeptide from strain SF162; SEQ ID NO:88, which depicts a tat construct encoding a tat polypeptide having the cysteine residue at position 22 changed; and SEQ ID NO:89, which depicts a tat construct 20 encoding the amino terminal portion of a tat polypeptide from strain SF162. The amino portion of the tat protein appears to contain many of the epitopes that induce an immune response. In addition, further modifications include replacement or deletion of the cysteine residue at 25 position 22, for example with a valine residue, an alanine residue or a glycine residue (SEQ ID Nos: 88 and 89, Figures 79 and 81), see, e.g., Caputo et al. (1996) Gene Ther. 3:235. In Figure 81, which depicts a tat construct encoding the amino terminal portion of a tat 30 polypeptide, the nucleotides (nucleotides 64-66) encoding the cysteine residues are underlined. The design and construction of suitable construct can be readily done using

the teachings of the present specification. As with Gag, pol, prot and Env, tat polypeptide coding sequences can be obtained from a variety of isolates (families, subtypes, etc.).

5 Modification of the tat polypeptide coding sequences result in improved expression relative to the wild-type coding sequences in a number of cell lines (e.g., mammalian, yeast, bacterial and insect cells). Tat polypeptide encoding sequences derived from these
10 variants can be optimized and tested for improved expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 2).

15 Various forms of the different embodiments of the invention, described herein, may be combined. For example, polynucleotides may be derived from the polynucleotide sequences of the present invention, including, but not limited to, coding sequences for Gag polypeptides, Env polypeptides, polymerase polypeptides,
20 protease polypeptides, tat polypeptides, and reverse transcriptase polypeptides. Further, the polynucleotide coding sequences of the present invention may be combined into multi-cistronic expression cassettes where typically each coding sequence for each polypeptide is preceded by
25 IRES sequences.

2.2 PRODUCTION OF VIRUS-LIKE PARTICLES AND USE OF THE CONSTRUCTS OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

30 The group-specific antigens (Gag) of human immunodeficiency virus type-1 (HIV-1) self-assemble into noninfectious virus-like particles (VLP) that are released from various eucaryotic cells by budding (reviewed by Freed, E.O., *Virology* 251:1-15, 1998). The

synthetic expression cassettes of the present invention provide efficient means for the production of HIV-Gag virus-like particles (VLPs) using a variety of different cell types, including, but not limited to, mammalian cells.

5 Viral particles can be used as a matrix for the proper presentation of an antigen entrapped or associated therewith to the immune system of the host. For example, U.S. Patent No. 4,722,840 describes hybrid particles 10 comprised of a particle-forming fragment of a structural protein from a virus, such as a particle-forming fragment of hepatitis B virus (HBV) surface antigen (HBsAg), fused to a heterologous polypeptide. Tindle et al., *Virology* (1994) 200:547-557, describes the production and use of 15 chimeric HBV core antigen particles containing epitopes of human papillomavirus (HPV) type 16 E7 transforming protein.

Adams et al., *Nature* (1987) 329:68-70, describes the recombinant production of hybrid HIVgp120:Ty VLPs in 20 yeast and Brown et al., *Virology* (1994) 198:477-488, the production of chimeric proteins consisting of the VP2 protein of human parvovirus B19 and epitopes from human herpes simplex virus type 1, as well as mouse hepatitis virus A59. Wagner et al., (*Virology* (1994) 200:162-175, 25 Brand et al., *J. Virol. Meth.* (1995) 51:153-168; *Virology* (1996) 220:128-140) and Wolf, et al., (EP 0 449 116 A1, published 2 October 1991; WO 96/30523, published 3 October 1996) describe the assembly of chimeric HIV-1 p55Gag particles. U.S. Patent No. 5,503,833 describes 30 the use of rotavirus VP6 spheres for encapsulating and delivering therapeutic agents.

2.2.1 VLP PRODUCTION USING THE SYNTHETIC EXPRESSION
CASSETTES OF THE PRESENT INVENTION

Experiments performed in support of the present invention have demonstrated that the synthetic expression cassettes of the present invention provide superior production of both protein and VLPs, relative to native coding sequences (Examples 7 and 15). Further, electron microscopic evaluation of VLP production (Examples 6 and 15, Figures 3A-B and 65A-F) showed that free and budding immature virus particles of the expected size were produced by cells containing the synthetic expression cassettes.

Using the synthetic expression cassettes of the present invention, rather than native coding sequences, for the production of virus-like particles provide several advantages. First, VLPs can be produced in enhanced quantity making isolation and purification of the VLPs easier. Second, VLPs can be produced in a variety of cell types using the synthetic expression cassettes, in particular, mammalian cell lines can be used for VLP production, for example, CHO cells. Production using CHO cells provides (i) VLP formation; (ii) correct myristylation and budding; (iii) absence of non-mammalian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification. The synthetic expression cassettes of the present invention are also useful for enhanced expression in cell-types other than mammalian cell lines. For example, infection of insect cells with baculovirus vectors encoding the synthetic expression cassettes resulted in higher levels of total protein yield and higher levels of VLP production (relative to wild-type coding sequences). Further, the final product from insect cells infected with the baculovirus-Gag synthetic expression cassettes

consistently contained lower amounts of contaminating insect proteins than the final product when wild-type coding sequences were used (Examples).

VLPs can spontaneously form when the particle-forming polypeptide of interest is recombinantly expressed in an appropriate host cell. Thus, the VLPs produced using the synthetic expression cassettes of the present invention are conveniently prepared using recombinant techniques. As discussed below, the Gag polypeptide encoding synthetic expression cassettes of the present invention can include other polypeptide coding sequences of interest (for example, Env, tat, rev, HIV protease, HIV polymerase, HCV core; see, Example 1). Expression of such synthetic expression cassettes yields VLPs comprising the product of the synthetic expression cassette, as well as, the polypeptide of interest.

Once coding sequences for the desired particle-forming polypeptides have been isolated or synthesized, they can be cloned into any suitable vector or replicon for expression. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. See, generally, Ausubel et al, *supra* or Sambrook et al, *supra*. The vector is then used to transform an appropriate host cell. Suitable recombinant expression systems include, but are not limited to, bacterial, mammalian, baculovirus/insect, vaccinia, Semliki Forest virus (SFV), Alphaviruses (such as, Sindbis, Venezuelan Equine Encephalitis (VEE)), mammalian, yeast and *Xenopus* expression systems, well known in the art. Particularly preferred expression systems are mammalian cell lines, vaccinia, Sindbis, insect and yeast systems.

For example, a number of mammalian cell lines are known in the art and include immortalized cell lines

available from the American Type Culture Collection (A.T.C.C.), such as, but not limited to, Chinese hamster ovary (CHO) cells, 293 cells, HeLa cells, baby hamster kidney (BHK) cells, mouse myeloma (SB20), monkey kidney cells (COS), as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*, will find use with the present expression constructs. Yeast hosts useful in the present invention include *inter alia*, *Saccharomyces cerevisiae*, 5 *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillermondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*. See, e.g., Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987). Fungal hosts include, for example, *Aspergillus*.

20 Viral vectors can be used for the production of particles in eucaryotic cells, such as those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. Additionally, a vaccinia based infection/transfection system, as described in Tomei et al., *J. Virol.* (1993) 67:4017-4026 and Selby et al., *J. Gen. Virol.* (1993) 74:1103-1113, will also find use with the present invention. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This 25 polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus

recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. Alternately, T7 can be added as a purified protein or enzyme as in the "Progenitor" system (Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130). The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation product(s).

Depending on the expression system and host selected, the VLPs are produced by growing host cells transformed by an expression vector under conditions whereby the particle-forming polypeptide is expressed and VLPs can be formed. The selection of the appropriate growth conditions is within the skill of the art. If the VLPs are formed intracellularly, the cells are then disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the VLPs substantially intact. Such methods are known to those of skill in the art and are described in, e.g., *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990).

The particles are then isolated (or substantially purified) using methods that preserve the integrity thereof, such as, by density gradient centrifugation, e.g., sucrose gradients, PEG-precipitation, pelleting, and the like (see, e.g., Kirnbauer et al. *J. Virol.* (1993) 67:6929-6936), as well as standard purification techniques including, e.g., ion exchange and gel filtration chromatography.

VLPs produced by cells containing the synthetic expression cassettes of the present invention can be used to elicit an immune response when administered to a subject. One advantage of the present invention is that VLPs can be produced by mammalian cells carrying the

synthetic expression cassettes at levels previously not possible. As discussed above, the VLPs can comprise a variety of antigens in addition to the Gag polypeptides (e.g., Env, tat, Gag-protease, Gag-polymerase, Gag-HCV- core). Purified VLPs, produced using the synthetic expression cassettes of the present invention, can be administered to a vertebrate subject, usually in the form of vaccine compositions. Combination vaccines may also be used, where such vaccines contain, for example, other 5 subunit proteins derived from HIV or other organisms (e.g., env) or gene delivery vaccines encoding such 10 antigens. Administration can take place using the VLPs formulated alone or formulated with other antigens. Further, the VLPs can be administered prior to, 15 concurrent with, or subsequent to, delivery of the synthetic expression cassettes for DNA immunization (see below) and/or delivery of other vaccines. Also, the site of VLP administration may be the same or different as other vaccine compositions that are being administered. 20 Gene delivery can be accomplished by a number of methods including, but are not limited to, immunization with DNA, alphavirus vectors, pox virus vectors, and vaccinia virus vectors.

VLP immune-stimulating (or vaccine) compositions can 25 include various excipients, adjuvants, carriers, auxiliary substances, modulating agents, and the like. The immune stimulating compositions will include an amount of the VLP/antigen sufficient to mount an immunological response. An appropriate effective amount 30 can be determined by one of skill in the art. Such an amount will fall in a relatively broad range that can be determined through routine trials and will generally be an amount on the order of about 0.1 µg to about 1000 µg,

more preferably about 1 μ g to about 300 μ g, of VLP/antigen.

A carrier is optionally present which is a molecule that does not itself induce the production of antibodies harmful to the individual receiving the composition.

Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc., as well as toxins derived from *E. coli*.

Such adjuvants include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (International Publication No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated

into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) 5 either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group 10 consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particle generated therefrom such as 15 ISCOMs (immunostimulating complexes); (4) Complete Freunds Adjuvant (CFA) and Incomplete Freunds Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), beta chemokines (MIP, 1- 20 alpha, 1-beta Rantes, etc.); (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63) 25 LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 (where lysine is substituted for the wild-type amino acid at position 9 and glycine substituted at position 129) (see, e.g., International 30 Publication Nos. W093/13202 and W092/19265); and (7)

other substances that act as immunostimulating agents to enhance the effectiveness of the composition.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isogluatme (nor-MDP), N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

Dosage treatment with the VLP composition may be a single dose schedule or a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals, chosen to maintain and/or reinforce the immune response, for example at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the potency of the modality, the vaccine delivery employed, the need of the subject and be dependent on the judgment of the practitioner.

If prevention of disease is desired (e.g., reduction of symptoms, recurrences or of disease progression), the antigen carrying VLPs are generally administered prior to primary infection with the pathogen of interest. If treatment is desired, e.g., the reduction of symptoms or recurrences, the VLP compositions are generally administered subsequent to primary infection.

2.2.2 USING THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

A number of viral based systems have been developed for use as gene transfer vectors for mammalian host cells. For example, retroviruses (in particular,

lentiviral vectors) provide a convenient platform for gene delivery systems. A coding sequence of interest (for example, a sequence useful for gene therapy applications) can be inserted into a gene delivery vector and packaged in retroviral particles using techniques known in the art. Recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described, including, for example, the following: (U.S. Patent No. 5,219,740; Miller et al. (1989) *Biotechniques* 7:980; Miller, A.D. (1990) *Human Gene Therapy* 1:5; Scarpa et al. (1991) *Virology* 180:849; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033; Boris-Lawrie et al. (1993) *Cur. Opin. Genet. Develop.* 3:102; GB 2200651; EP 0415731; EP 0345242; WO 89/02468; WO 89/05349; WO 89/09271; WO 90/02806; WO 90/07936; WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; WO 93/11230; WO 93/10218; WO 91/02805; in U.S. 5,219,740; U.S. 4,405,712; U.S. 4,861,719; U.S. 4,980,289 and U.S. 4,777,127; in U.S. Serial No. 07/800,921; and in Vile (1993) *Cancer Res.* 53:3860-3864; Vile (1993) *Cancer Res.* 53:962-967; Ram (1993) *Cancer Res.* 53:83-88; Takamiya (1992) *J Neurosci.* 12:493-503; Baba (1993) *J Neurosurg.* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci USA* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Sequences useful for gene therapy applications include, but are not limited to, the following. Factor VIII cDNA, including derivatives and deletions thereof (International Publication Nos. WO 96/21035, WO 97/03193, WO 97/03194, WO 97/03195, and WO 97/03191). Factor IX cDNA (Kurachi et al. (1982) *Proc. Natl. Acad. Sci. USA* 79:6461-6464). Factor V cDNA can be obtained from pMT2-V (Jenny (1987) *Proc. Natl. Acad. Sci. USA* 84:4846, A.T.C.C. Deposit No. 40515). A full-length factor V

cDNA, or a B domain deletion or B domain substitution thereof, can be used. B domain deletions of factor V, include those reported by Marquette (1995) *Blood* 86:3026 and Kane (1990) *Biochemistry* 29:6762. Antithrombin III 5 cDNA (Prochownik (1983) *J. Biol. Chem.* 258:8389, A.T.C.C. Deposit No. 57224/57225). Protein C encoding cDNA (Foster (1984) *Proc. Natl. Acad. Sci. USA* 81:4766; Beckmann (1985) *Nucleic Acids Res.* 13:5233). Prothrombin cDNA can be obtained by restriction enzyme digestion of a 10 published vector (Degen (1983) *Biochemistry* 22:2087). The endothelial cell surface protein, thrombomodulin, is a necessary cofactor for the normal activation of protein C by thrombin. A soluble recombinant form has been described (Parkinson (1990) *J. Biol. Chem.* 265:12602; 15 Jackman (1987) *Proc. Natl. Acad. Sci. USA* 84:6425; Shirai (1988) *J. Biochem.* 103:281; Wen (1987) *Biochemistry* 26:4350; Suzuki (1987) *EMBO J.* 6:1891, A.T.C.C. Deposit No. 61348, 61349).

Many genetic diseases caused by inheritance of 20 defective genes result in the failure to produce normal gene products, for example, thalassemia, phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency (SCID), hemophilia A and B, cystic fibrosis, Duchenne's Muscular Dystrophy, inherited emphysema and familial 25 hypercholesterolemia (Mulligan et al. (1993) *Science* 260:926; Anderson et al. (1992) *Science* 256:808; Friedman et al. (1989) *Science* 244:1275). Although genetic diseases may result in the absence of a gene product, endocrine disorders, such as diabetes and 30 hypopituitarism, are caused by the inability of the gene to produce adequate levels of the appropriate hormone insulin and human growth hormone respectively.

In one aspect, gene therapy employing the constructs and methods of the present invention involves the

introduction of normal recombinant genes into T cells so that new or missing proteins are produced by the T cells after introduction or reintroduction thereof into a patient. A number of genetic diseases have been selected 5 for treatment with gene therapy, including adenine deaminase deficiency, cystic fibrosis, α_1 -antitrypsin deficiency, Gaucher's syndrome, as well as non-genetic diseases.

In particular, Gaucher's syndrome is a genetic disorder characterized by a deficiency of the enzyme 10 glucocerebrosidase. This enzyme deficiency leads to the accumulation of glucocerebroside in the lysosomes of all cells in the body. For a review see *Science* 256:794 (1992) and Scriver et al., *The Metabolic Basis of 15 Inherited Disease*, 6th ed., vol. 2, page 1677). Thus, gene transfer vectors that express glucocerebrosidase can be constructed for use in the treatment of this disorder. Likewise, gene transfer vectors encoding lactase can be used in the treatment of hereditary lactose intolerance, 20 those expressing ADA can be used for treatment of ADA deficiency, and gene transfer vectors encoding α_1 -antitrypsin can be used to treat α_1 -antitrypsin deficiency. See Ledley, F.D. (1987) *J. Pediatrics* 110:157-174, Verma, I. (Nov. 1987) *Scientific American* 25 pp. 68-84, and International Publication No. WO 95/27512 entitled "Gene Therapy Treatment for a Variety of Diseases and Disorders," for a description of gene 25 therapy treatment of genetic diseases.

In still further embodiments of the invention, 30 nucleotide sequences which can be incorporated into a gene transfer vector include, but are not limited to, proteins associated with enzyme-deficiency disorders, such as the cystic fibrosis transmembrane regulator (see, for example, U.S. Patent No. 5,240,846 and Larrick et al.

(1991) *Gene Therapy Applications of Molecular Biology*, Elsevier, New York and adenosine deaminase (ADA) (see U.S. Patent No. 5,399,346); growth factors, or an agonist or antagonist of a growth factor (Bandara et al. (1992) 5 *DNA and Cell Biology*, 11:227); one or more tumor suppressor genes such as p53, Rb, or C-CAMI (Kleinerman et al. (1995) *Cancer Research* 55:2831); a molecule that modulates the immune system of an organism, such as a HLA molecule (Nabel et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:11307); a ribozyme (Larsson et al. (1996) *Virology* 219:161); a peptide nucleic acid (Hirshman et al. (1996) *J. Invest. Med.* 44:347); an antisense molecule (Bordier et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:9383) which can be used to down-regulate the expression or synthesis 10 of aberrant or foreign proteins, such as HIV proteins or 15 a wide variety of oncogenes such as p53 (Hesketh, *The Oncogene Facts Book*, Academic Press, New York, (1995); a biopharmaceutical agent or antisense molecule used to treat HIV-infection, such as an inhibitor of p24 20 (Nakashima et al. (1994) *Nucleic Acids Res.* 22:5004); or reverse-transcriptase (see, Bordier, *supra*).

Other proteins of therapeutic interest can be expressed *in vivo* by gene transfer vectors using the methods of the invention. For instance sustained *in vivo* expression of tissue factor inhibitory protein (TFPI) is useful for treatment of conditions including sepsis and DIC and in preventing reperfusion injury. (See 25 International Publications Nos. WO 93/24143, WO 93/25230 and WO 96/06637). Nucleic acid sequences encoding expression of TFPI can be obtained, for example, as 30 described in US Patent Nos. 4,966,852; 5,106,833; and 5,466,783, and incorporated into the gene transfer vectors described herein.

Erythropoietin (EPO) and leptin can also be expressed in vivo from genetically modified T cells according to the methods of the invention. For instance EPO is useful in gene therapy treatment of a variety of disorders including anemia (see International Publication No. WO 95/13376 entitled "Gene Therapy for Treatment of Anemia"). Sustained delivery of leptin by the methods of the invention is useful in treatment of obesity. See International Publication No. WO 96/05309 for a description of the leptin gene and the use thereof in the treatment of obesity.

A variety of other disorders can also be treated by the methods of the invention. For example, sustained in vivo systemic production of apolipoprotein E or apolipoprotein A from genetically modified T cells can be used for treatment of hyperlipidemia (see Breslow et al. (1994) *Biotechnology* 12:365). Sustained production of angiotensin receptor inhibitor (Goodfriend et al. (1996) *N. Engl. J. Med.* 334:1469) can be provided by the methods described herein. As yet an additional example, the long term in vivo systemic production of angiostatin is useful in the treatment of a variety of tumors. (See O'Reilly et al. (1996) *Nature Med.* 2:689).

In other embodiments, gene transfer vectors can be constructed to encode a cytokine or other immunomodulatory molecule. For example, nucleic acid sequences encoding native IL-2 and gamma-interferon can be obtained as described in US Patent Nos. 4,738,927 and 5,326,859, respectively, while useful mutants of these proteins can be obtained as described in U.S. Patent No. 4,853,332. Nucleic acid sequences encoding the short and long forms of mCSF can be obtained as described in US Patent Nos. 4,847,201 and 4,879,227, respectively. In particular aspects of the invention, retroviral vectors

expressing cytokine or immunomodulatory genes can be produced as described herein (for example, employing the packaging cell lines of the present invention) and in International Application No. PCT US 94/02951, entitled 5 "Compositions and Methods for Cancer Immunotherapy."

Examples of suitable immunomodulatory molecules for use herein include the following: IL-1 and IL-2 (Karupiah et al. (1990) *J. Immunology* 144:290-298, Weber et al. (1987) *J. Exp. Med.* 166:1716-1733, Gansbacher et al. 10 (1990) *J. Exp. Med.* 172:1217-1224, and U.S. Patent No. 4,738,927); IL-3 and IL-4 (Tepper et al. (1989) *Cell* 57:503-512, Columbek et al. (1991) *Science* 254:713-716, and U.S. Patent No. 5,017,691); IL-5 and IL-6 (Brakenhof et al. (1987) *J. Immunol.* 139:4116-4121, and 15 International Publication No. WO 90/06370); IL-7 (U.S. Patent No. 4,965,195); IL-8, IL-9, IL-10, IL-11, IL-12, and IL-13 (*Cytokine Bulletin*, Summer 1994); IL-14 and IL-15; alpha interferon (Finter et al. (1991) *Drugs* 42:749-765, U.S. Patent Nos. 4,892,743 and 4,966,843, 20 International Publication No. WO 85/02862, Nagata et al. (1980) *Nature* 284:316-320, Familletti et al. (1981) *Methods in Enz.* 78:387-394, Twu et al. (1989) *Proc. Nati. Acad. Sci. USA* 86:2046-2050, and Faktor et al. (1990) *Oncogene* 5:867-872); beta-interferon (Seif et al. (1991) *J. Virol.* 65:664-671); gamma-interferons (Radford et al. 25 (1991) *The American Society of Hepatology* 20082015, Watanabe et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:9456-9460, Gansbacher et al. (1990) *Cancer Research* 50:7820-7825, Maio et al. (1989) *Can. Immunol.* 30 *Immunother.* 30:34-42, and U.S. Patent Nos. 4,762,791 and 4,727,138); G-CSF (U.S. Patent Nos. 4,999,291 and 4,810,643); GM-CSF (International Publication No. WO 85/04188); tumor necrosis factors (TNFs) (Jayaraman et al. (1990) *J. Immunology* 144:942-951); CD3 (Krissanen et

al. (1987) *Immunogenetics* 26:258-266); ICAM-1 (Altman et al. (1989) *Nature* 338:512-514, Simmons et al. (1988) *Nature* 331:624-627); ICAM-2, LFA-1, LFA-3 (Wallner et al. (1987) *J. Exp. Med.* 166:923-932); MHC class I molecules, 5 MHC class II molecules, B7.1-.3, β_2 -microglobulin (Parnes et al. (1981) *Proc. Natl. Acad. Sci. USA* 78:2253-2257); chaperones such as calnexin; and MHC-linked transporter proteins or analogs thereof (Powis et al. (1991) *Nature* 354:528-531). Immunomodulatory factors may also be 10 agonists, antagonists, or ligands for these molecules. For example, soluble forms of receptors can often behave as antagonists for these types of factors, as can mutated forms of the factors themselves.

Nucleic acid molecules that encode the above-described substances, as well as other nucleic acid molecules that are advantageous for use within the present invention, may be readily obtained from a variety of sources, including, for example, depositories such as the American Type Culture Collection, or from commercial 15 sources such as British Bio-Technology Limited (Cowley, Oxford England). Representative examples include BBG 12 (containing the GM-CSF gene coding for the mature protein of 127 amino acids), BBG 6 (which contains sequences encoding gamma interferon), A.T.C.C. Deposit No. 39656 20 (which contains sequences encoding TNF), A.T.C.C. Deposit No. 20663 (which contains sequences encoding alpha-interferon), A.T.C.C. Deposit Nos. 31902, 31902 and 39517 (which contain sequences encoding beta-interferon), A.T.C.C. Deposit No. 67024 (which contains a sequence 25 which encodes Interleukin-1b), A.T.C.C. Deposit Nos. 39405, 39452, 39516, 39626 and 39673 (which contain sequences encoding Interleukin-2), A.T.C.C. Deposit Nos. 59399, 59398, and 67326 (which contain sequences encoding 30 Interleukin-3), A.T.C.C. Deposit No. 57592 (which

contains sequences encoding Interleukin-4), A.T.C.C. Deposit Nos. 59394 and 59395 (which contain sequences encoding Interleukin-5), and A.T.C.C. Deposit No. 67153 (which contains sequences encoding Interleukin-6).

5 Plasmids containing cytokine genes or immunomodulatory genes (International Publication Nos. WO 94/02951 and WO 96/21015) can be digested with appropriate restriction enzymes, and DNA fragments containing the particular gene of interest can be inserted into a gene transfer vector using standard molecular biology techniques. (See, e.g., Sambrook et al., *supra.*, or Ausubel et al. (eds) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience).

10 Exemplary hormones, growth factors and other proteins which are useful for long term expression are described, for example, in European Publication No. 0437478B1, entitled "Cyclodextrin-Peptide Complexes." Nucleic acid sequences encoding a variety of hormones can be used, including those encoding human growth hormone, insulin, calcitonin, prolactin, follicle stimulating hormone (FSH), luteinizing hormone (LH), human chorionic gonadotropin (HCG), and thyroid stimulating hormone (TSH). A variety of different forms of IGF-1 and IGF-2 growth factor polypeptides are also well known the art and can be incorporated into gene transfer vectors for long term expression *in vivo*. See, e.g., European Patent No. 0123228B1, published for grant September 19, 1993, entitled "Hybrid DNA Synthesis of Mature Insulin-like Growth Factors." As an additional example, the long term 15 *in vivo* expression of different forms of fibroblast growth factor can also be effected employing the compositions and methods of invention. See, e.g., U.S. Patent Nos. 5,464,774, 5,155,214, and 4,994,559 for a description of different fibroblast growth factors.

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene 5 from a vector known to include the same. For example, plasmids which contain sequences that encode altered cellular products may be obtained from a depository such as the A.T.C.C., or from commercial sources. Plasmids containing the nucleotide sequences of interest can be 10 digested with appropriate restriction enzymes, and DNA fragments containing the nucleotide sequences can be inserted into a gene transfer vector using standard molecular biology techniques.

Alternatively, cDNA sequences for use with the 15 present invention may be obtained from cells which express or contain the sequences, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA. 20 Briefly, mRNA from a cell which expresses the gene of interest can be reverse transcribed with reverse transcriptase using oligo-dT or random primers. The single stranded cDNA may then be amplified by PCR (see U.S. Patent Nos. 4,683,202, 4,683,195 and 4,800,159, see 25 also *PCR Technology: Principles and Applications for DNA Amplification*, Erlich (ed.), Stockton Press, 1989)) using oligonucleotide primers complementary to sequences on either side of desired sequences.

The nucleotide sequence of interest can also be 30 produced synthetically, rather than cloned, using a DNA synthesizer (e.g., an Applied Biosystems Model 392 DNA Synthesizer, available from ABI, Foster City, California). The nucleotide sequence can be designed with the appropriate codons for the expression product

desired. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge (1981) *Nature* 292:756; Nambair et al. (1984) Science 223:1299; Jay et al. (1984) *J. Biol. Chem.* 259:6311.

5 The synthetic expression cassettes of the present invention can be employed in the construction of packaging cell lines for use with retroviral vectors.

10 One type of retrovirus, the murine leukemia virus, or "MLV", has been widely utilized for gene therapy applications (see generally Mann et al. (*Cell* 33:153, 1993), Cane and Mulligan. (*Proc, Nat'l. Acad. Sci. USA* 81:6349, 1984), and Miller et al., *Human Gene Therapy* 1:5-14, 1990.

15 Lentiviral vectors typically, comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, a promoter operably linked to one or more genes of interest, an origin of second strand DNA synthesis and a 20 3' lentiviral LTR, wherein the lentiviral vector contains a nuclear transport element. The nuclear transport element may be located either upstream (5') or downstream (3') of a coding sequence of interest. Within certain embodiments, the nuclear transport element is not RRE. 25 Within one embodiment the packaging signal is an extended packaging signal. Within other embodiments the promoter is a tissue specific promoter, or, alternatively, a promoter such as CMV. Within other embodiments, the lentiviral vector further comprises an internal ribosome 30 entry site.

A wide variety of lentiviruses may be utilized within the context of the present invention, including for example, lentiviruses selected from the group consisting of HIV, HIV-1, HIV-2, FIV and SIV.

In one embodiment of the present invention synthetic Env and/or Gag-polymerase expression cassettes are provided comprising a promoter and a sequence encoding synthetic Gag-polymerase (SEQ ID NO:6) and at least one of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif.

Within yet another aspect of the invention, host cells (e.g., packaging cell lines) are provided which contain any of the expression cassettes described herein.

For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Env and/or Gag-polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Env and/or Gag-polymerase. Packaging cell lines may further comprise a promoter and a sequence encoding tat, rev, or an envelope, wherein the promoter is operably linked to the sequence encoding tat, rev, or, the envelope. The packaging cell line may further comprise a sequence encoding any one or more of nef, vif, vpu or vpr.

In one embodiment, the expression cassette (carrying, for example, the synthetic Env, synthetic tat and/or synthetic Gag-polymerase) is stably integrated. The packaging cell line, upon introduction of a lentiviral vector, typically produces viral particles. The promoter regulating expression of the synthetic expression cassette may be inducible. Typically, the packaging cell line, upon introduction of a lentiviral vector, produces viral particles that are essentially free of replication competent virus.

Packaging cell lines are provided comprising an expression cassette which directs the expression of a synthetic Env (or Gag-polymerase) gene, an expression cassette which directs the expression of a Gag (or Env)

gene optimized for expression (e.g., Andre, S., et al., *Journal of Virology* 72(2):1497-1503, 1998; Haas, J., et al., *Current Biology* 6(3):315-324, 1996). A lentiviral vector is introduced into the packaging cell line to produce a vector particle producing cell line.

As noted above, lentiviral vectors can be designed to carry or express a selected gene(s) or sequences of interest. Lentiviral vectors may be readily constructed from a wide variety of lentiviruses (see RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985). Representative examples of lentiviruses included HIV, HIV-1, HIV-2, FIV and SIV. Such lentiviruses may either be obtained from patient isolates, or, more preferably, from depositories or collections such as the American Type Culture Collection, or isolated from known sources using available techniques.

Portions of the lentiviral gene delivery vectors (or vehicles) may be derived from different viruses. For example, in a given recombinant lentiviral vector, LTRs may be derived from an HIV, a packaging signal from SIV, and an origin of second strand synthesis from HrV-2. Lentiviral vector constructs may comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, one or more heterologous sequences, an origin of second strand DNA synthesis and a 3' LTR, wherein said lentiviral vector contains a nuclear transport element that is not RRE.

Briefly, Long Terminal Repeats ("LTRs") are subdivided into three elements, designated U5, R and U3. These elements contain a variety of signals which are responsible for the biological activity of a retrovirus, including for example, promoter and enhancer elements which are located within U3. LTRs may be readily identified in the provirus (integrated DNA form) due to their precise duplication at either end of the genome.

As utilized herein, a 5' LTR should be understood to include a 5' promoter element and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector. The 3' LTR should be understood to 5 include a polyadenylation signal, and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector.

The tRNA binding site and origin of second strand DNA synthesis are also important for a retrovirus to be 10 biologically active, and may be readily identified by one of skill in the art. For example, retroviral tRNA binds to a tRNA binding site by Watson-Crick base pairing, and is carried with the retrovirus genome into a viral 15 particle. The tRNA is then utilized as a primer for DNA synthesis by reverse transcriptase. The tRNA binding site may be readily identified based upon its location just downstream from the 5'LTR. Similarly, the origin of second strand DNA synthesis is, as its name implies, 20 important for the second strand DNA synthesis of a retrovirus. This region, which is also referred to as the poly-purine tract, is located just upstream of the 3'LTR.

In addition to a 5' and 3' LTR, tRNA binding site, and origin of second strand DNA synthesis, recombinant 25 retroviral vector constructs may also comprise a packaging signal, as well as one or more genes or coding sequences of interest. In addition, the lentiviral vectors have a nuclear transport element which, in preferred embodiments is not RRE. Representative 30 examples of suitable nuclear transport elements include the element in Rous sarcoma virus (Ogert, et al., J ViroL 70, 3834-3843, 1996), the element in Rous sarcoma virus (Liu & Mertz, Genes & Dev., 9, 1766-1789, 1995) and the element in the genome of simian retrovirus type I

WO 00/39302

(Zolotukhin, et al., *J Virol.* 68, 7944-7952, 1994). Other potential elements include the elements in the histone gene (Kedes, *Annu. Rev. Biochem.* 48, 837-870, 1970), the α -interferon gene (Nagata et al., *Nature* 287, 5 401-408, 1980), the β -adrenergic receptor gene (Koilka, et al., *Nature* 329, 75-79, 1987), and the c-Jun gene (Hattorie, et al., *Proc. Natl. Acad. Sci. USA* 85, 9148-9152, 1988).

Recombinant lentiviral vector constructs typically lack both *Gag*-polymerase and *env* coding sequences. Recombinant lentiviral vector typically contain less than 10, preferably 15, more preferably 10, and most preferably 8 consecutive nucleotides found in *Gag*-polymerase or *env* genes. One advantage of the present invention is that the synthetic *Gag*-polymerase expression cassettes, which can be used to construct packaging cell lines for the recombinant retroviral vector constructs, have little homology to wild-type *Gag*-polymerase sequences and thus considerably reduce or eliminate the 15 possibility of homologous recombination between the 20 synthetic and wild-type sequences.

Lentiviral vectors may also include tissue-specific promoters to drive expression of one or more genes or sequences of interest. For example, lentiviral vector particles of the invention can contain a liver specific promoter to maximize the potential for liver specific expression of the exogenous DNA sequence contained in the vectors. Preferred liver specific promoters include the hepatitis B X-gene promoter and the hepatitis B core protein promoter. These liver specific promoters are 25 preferably employed with their respective enhancers. The enhancer element can be linked at either the 5' or the 3' 30 end of the nucleic acid encoding the sequences of interest. The hepatitis B X gene promoter and its

enhancer can be obtained from the viral genome as a 332 base pair EcoRV-NcoI DNA fragment employing the methods described in Twu, et al., *J Virol.* 61:3448-3453, 1987. The hepatitis B core protein promoter can be obtained 5 from the viral genome as a 584 base pair BamHI-BglII DNA fragment employing the methods described in Gerlach, et al., *Virol* 189:59-66, 1992. It may be necessary to remove the negative regulatory sequence in the BamHI-BglII fragment prior to inserting it. Other liver 10 specific promoters include the AFP (alpha fetal protein) gene promoter and the albumin gene promoter, as disclosed in EP Patent Publication 0 415 731, the -1 antitrypsin gene promoter, as disclosed in Rettenger, et al., *Proc. Natl. Acad. Sci.* 91:1460-1464, 1994, the fibrinogen 15 gene promoter, the APO-A1 (Apolipoprotein A1) gene promoter, and the promoter genes for liver transference enzymes such as, for example, SGOT, SGPT and glutamyle transferase. See also PCT Patent Publications WO 90/07936 and WO 91/02805 for a description of the use of 20 liver specific promoters in lentiviral vector particles.

Lentiviral vector constructs may be generated such that more than one gene of interest is expressed. This may be accomplished through the use of di- or oligo-cistronic cassettes (e.g., where the coding regions are 25 separated by 80 nucleotides or less, see generally Levin et al., *Gene* 108:167-174, 1991), or through the use of Internal Ribosome Entry Sites ("IRES").

Packaging cell lines suitable for use with the above described recombinant retroviral vector constructs may be 30 readily prepared given the disclosure provided herein. Briefly, the parent cell line from which the packaging cell line is derived can be selected from a variety of

mammalian cell lines, including for example, 293, RD, COS-7, CHO, BHK, VERO, HT1080, and myeloma cells.

After selection of a suitable host cell for the generation of a packaging cell line, one or more expression cassettes are introduced into the cell line in order to complement or supply in *trans* components of the vector which have been deleted.

Representative examples of suitable expression cassettes have been described herein and include synthetic Env, tat, Gag, synthetic Gag-protease, synthetic Gag-reverse transcriptase and synthetic Gag-polymerase expression cassettes, which comprise a promoter and a sequence encoding, e.g., Env, tat, or Gag-polymerase and at least one of vpr, vpu, nef or vif, wherein the promoter is operably linked to Env, tat or Gag-polymerase and vpr, vpu, nef or vif. As described above, optimized Env, Gag and/or tat coding sequences may also be utilized in various combinations in the generation of packaging cell lines.

Utilizing the above-described expression cassettes, a wide variety of packaging cell lines can be generated. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic HIV (e.g., Gag, Env, tat, Gag-polymerase, Gag-reverse transcriptase or Gag-protease) polypeptide, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding the HIV polypeptide. Within other aspects, packaging cell lines are provided comprising a promoter and a sequence encoding Gag, tat, rev, or an envelope (e.g., HIV env), wherein the promoter is operably linked to the sequence encoding Gag, tat, rev, or, the envelope. Within further embodiments, the packaging cell line may comprise a sequence encoding any one or more of nef, vif,

vpu or vpr. For example, the packaging cell line may contain only nef, vif, vpu, or vpr alone, nef and vif, nef and vpu, nef and vpr, vif and vpu, vif and vpr, vpu and vpr, nef vif and vpu, nef vif and vpr, nef vpu and vpr, vvir vpu and vpr, or, all four of nef vif vpu and vpr.

In one embodiment, the expression cassette is stably integrated. Within another embodiment, the packaging cell line, upon introduction of a lentiviral vector, produces particles. Within further embodiments the promoter is inducible. Within certain preferred embodiments of the invention, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are free of replication competent virus.

The synthetic cassettes containing optimized coding sequences are transfected into a selected cell line. Transfected cells are selected that (i) carry, typically, integrated, stable copies of the Gag, Pol, and Env coding sequences, and (ii) are expressing acceptable levels of these polypeptides (expression can be evaluated by methods known in the prior art, e.g., see Examples 1-4). The ability of the cell line to produce VLPs may also be verified (Examples 6, 7 and 15).

A sequence of interest is constructed into a suitable viral vector as discussed above. This defective virus is then transfected into the packaging cell line. The packaging cell line provides the viral functions necessary for producing virus-like particles into which the defective viral genome, containing the sequence of interest, are packaged. These VLPs are then isolated and can be used, for example, in gene delivery or gene therapy.

Further, such packaging cell lines can also be used to produce VLPs alone, which can, for example, be used as

adjuvants for administration with other antigens or in vaccine compositions. Also, co-expression of a selected sequence of interest encoding a polypeptide (for example, an antigen) in the packaging cell line can also result in 5 the entrapment and/or association of the selected polypeptide in/with the VLPs.

2.3 DNA IMMUNIZATION AND GENE DELIVERY

A variety of polypeptide antigens can be used in the 10 practice of the present invention. Polypeptide antigens can be included in DNA immunization constructs containing, for example, any of the synthetic expression cassettes described herein fused in-frame to a coding sequence for the polypeptide antigen, where expression of 15 the construct results in VLPs presenting the antigen of interest. Antigens can be derived from a wide variety of viruses, bacteria, fungi, plants, protozoans and other parasites. For example, the present invention will find use for stimulating an immune response against a wide 20 variety of proteins from the herpesvirus family, including proteins derived from herpes simplex virus (HSV) types 1 and 2, such as HSV-1 and HSV-2 gB, gD, gH, VP16 and VP22; antigens derived from varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus 25 (CMV) including CMV gB and gH; and antigens derived from other human herpesviruses such as HHV6 and HHV7. (See, e.g. Chee et al., *Cytomegaloviruses* (J.K. McDougall, ed., Springer-Verlag 1990) pp. 125-169, for a review of the protein coding content of cytomegalovirus; McGeoch et 30 al., *J. Gen. Virol.* (1988) 69:1531-1574, for a discussion of the various HSV-1 encoded proteins; U.S. Patent No. 5,171,568 for a discussion of HSV-1 and HSV-2 gB and gD proteins and the genes encoding therefore; Baer et al., *Nature* (1984) 310:207-211, for the identification of

WO 00/39302

protein coding sequences in an EBV genome; and Davison and Scott, *J. Gen. Virol.* (1986) 67:1759-1816, for a review of VZV.)

Additionally, immune responses to antigens from the hepatitis family of viruses, including hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV), and hepatitis G virus, can also be stimulated using the constructs of the present invention. By way of example, the HCV genome encodes several viral proteins, including E1 (also known as E) and E2 (also known as E2/NS1), which will find use with the present invention (see, Houghton et al. *Hepatology* (1991) 14:381-388, for a discussion of HCV proteins, including E1 and E2). The δ-antigen from HDV can also be used (see, e.g., U.S. Patent No. 5,389,528, for a description of the δ-antigen).

Similarly, influenza virus is another example of a virus for which the present invention will be particularly useful. Specifically, the envelope glycoproteins HA and NA of influenza A are of particular interest for generating an immune response. Numerous HA subtypes of influenza A have been identified (Kawaoka et al., *Virology* (1990) 179:759-767; Webster et al. "Antigenic variation among type A influenza viruses," p. 127-168. In: P. Palese and D.W. Kingsbury (ed.), *Genetics of influenza viruses*. Springer-Verlag, New York).

Other antigens of particular interest to be used in the practice of the present invention include antigens and polypeptides derived therefrom from human papillomavirus (HPV), such as one or more of the various early proteins including E6 and E7; tick-borne encephalitis viruses; and HIV-1 (also known as HTLV-III, LAV, ARV, etc.), including, but not limited to, antigens such as gp120, gp41, gp160, Gag and pol from a variety of

isolates including, but not limited to, HIV_{IIIB}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{us4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse 5 subtypes (e.g., HIV-2_{uc1} and HIV-2_{uc2}). See, e.g., Myers, et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico; Myers, et al., *Human Retroviruses and Aids*, 1990, Los Alamos, New Mexico: Los Alamos National Laboratory.

10 Proteins derived from other viruses will also find use in the claimed methods, such as without limitation, proteins from members of the families Picornaviridae (e.g., polioviruses, etc.); Caliciviridae; Togaviridae (e.g., rubella virus, dengue virus, etc.); Flaviviridae; 15 Coronaviridae; Reoviridae; Birnaviridae; Rhabdoviridae (e.g., rabies virus, etc.); Filoviridae; Paramyxoviridae (e.g., mumps virus, measles virus, respiratory syncytial virus, etc.); Orthomyxoviridae (e.g., influenza virus types A, B and C, etc.); Bunyaviridae; Arenaviridae; 20 Retroviridae, e.g., HTLV-I; HTLV-II; HIV-1; HIV-2; simian immunodeficiency virus (SIV) among others. See, e.g. *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM 25 Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA) for a description of these and other viruses.

Particularly preferred bacterial antigens are derived from organisms that cause diphtheria, tetanus, pertussis, meningitis, and other pathogenic states, 30 including, without limitation, antigens derived from *Corynebacterium diphtheriae*, *Clostridium tetani*, *Bordetella pertussis*, *Neisseria meningitidis*, including serotypes *Meningococcus A, B, C, Y and WI35* (MenA, B, C, Y and WI35), *Haemophilus influenza type B (Hib)*, and

Helicobacter pylori. Examples of parasitic antigens include those derived from organisms causing malaria, tuberculosis, and Lyme disease.

Furthermore, the methods described herein provide
5 means for treating a variety of malignant cancers. For example, the system of the present invention can be used to enhance both humoral and cell-mediated immune responses to particular proteins specific to a cancer in question, such as an activated oncogene, a fetal antigen,
10 or an activation marker. Such tumor antigens include any of the various MAGEs (melanoma associated antigen E), including MAGE 1, 2, 3, 4, etc. (Boon, T. *Scientific American* (March 1993):82-89); any of the various tyrosinases; MART 1 (melanoma antigen recognized by T
15 cells), mutant ras; mutant p53; p97 melanoma antigen; CEA (carcinoembryonic antigen), among others.

DNA immunization using synthetic expression cassettes of the present invention has been demonstrated to be efficacious (Examples 8 and 10-12). Animals were
20 immunized with both the synthetic expression cassette and the wild type expression cassette. The results of the immunizations with plasmid-DNAs showed that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression
25 cassettes. Also, the second boost immunization induced a secondary immune response, for example after two to eight weeks. Further, the results of CTL assays showed increased potency of synthetic expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by
30 DNA immunization.

It is readily apparent that the subject invention can be used to mount an immune response to a wide variety of antigens and hence to treat or prevent a large number of diseases.

2.3.1 DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE
PRESENT INVENTION

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. The sequences can be analyzed by conventional sequencing techniques. Furthermore, the desired gene can be isolated directly from cells and tissues containing the same, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain, isolate and sequence DNA. Once the sequence is known, the gene of interest can also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature* (1981) 292:756; Nambair et al., *Science* (1984) 223:1299; Jay et al., *J. Biol. Chem.* (1984) 259:6311; Stemmer, W.P.C., (1995) *Gene* 164:49-53.

Next, the gene sequence encoding the desired antigen can be inserted into a vector containing a synthetic expression cassette of the present invention (e.g., see Example 1 for construction of various exemplary synthetic expression cassette). The antigen is inserted into the synthetic coding sequence such that when the combined sequence is expressed it results in the production of VLPs comprising the polypeptide and/or the antigen of

WO 00/39302

interest. Insertions can be made within the Gag coding sequence or at either end of the coding sequence (5', amino terminus of the expressed polypeptide; or 3', carboxy terminus of the expressed polypeptide -- e.g., see Example 1) (Wagner, R., et al., Arch Virol. 127:117-137, 1992; Wagner, R., et al., Virology 200:162-175, 1994; Wu, X., et al., J. Virol. 69(6):3389-3398, 1995; Wang, C-T., et al., Virology 200:524-534, 1994; Chazal, N., et al., Virology 68(1):111-122, 1994; Griffiths, J.C., et al., J. Virol. 67(6):3191-3198, 1993; Reicin, A.S., et al., J. Virol. 69(2):642-650, 1995).

Up to 50% of the coding sequences of p55Gag can be deleted without affecting the assembly to virus-like particles and expression efficiency (Borsetti, A., et al., J. Virol. 72(11):9313-9317, 1998; Gamier, L., et al., J. Virol. 72(6):4667-4677, 1998; Zhang, Y., et al., J. Virol. 72(3):1782-1789, 1998; Wang, C., et al., J. Virol. 72(10):7950-7959, 1998). In one embodiment of the present invention, immunogenicity of the high level expressing synthetic p55GagMod and p55GagProtMod expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted, mutated or truncated regions of p55GagMod sequence. In another embodiment of the present invention, immunogenicity of the high level expressing synthetic Env expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted regions of gp120Mod, gp140Mod or gp160Mod sequences. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length modified Env

sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency and/or higher immunogenicity of the expression product. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length Env, Gag or Tat sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency and/or 10 immunogenicity of the expression product.

When sequences are added to the amino terminal end of Gag (for example, when using the synthetic p55GagMod expression cassette of the present invention), the polynucleotide can contain coding sequences at the 5' end 15 that encode a signal for addition of a myristic moiety to the Gag-containing polypeptide (e.g., sequences that encode Met-Gly).

The ability of Gag-containing polypeptide constructs to form VLPs can be empirically determined following the 20 teachings of the present specification.

HIV polypeptide/antigen synthetic expression cassettes include control elements operably linked to the coding sequence, which allow for the expression of the gene *in vivo* in the subject species. For example, 25 typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter, the mouse mammary tumor virus LTR promoter, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences 30 will also be present, located 3' to the translation stop

codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook et al., *supra*, as well as a bovine growth hormone terminator sequence.

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence.

Furthermore, plasmids can be constructed which include a chimeric antigen-coding gene sequences, encoding, e.g., multiple antigens/epitopes of interest, for example derived from a single or from more than one viral isolate.

Typically the antigen coding sequences precede or follow the synthetic coding sequences and the chimeric transcription unit will have a single open reading frame encoding both the antigen of interest and the synthetic Gag coding sequences. Alternatively, multi-cistronic cassettes (e.g., bi-cistronic cassettes) can be constructed allowing expression of multiple antigens from a single mRNA using the EMCV IRES, or the like. Lastly, antigens can be encoded on separate transcripts from independent promoters on a single plasmid or other vector.

Once complete, the constructs are used for nucleic acid immunization or the like using standard gene

delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, delivered *ex vivo*, to cells derived from the subject and the cells reimplanted in the subject.

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems. Selected sequences can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109.

A number of adenovirus vectors have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

Additionally, various adeno-associated virus (AAV) vector systems have been developed for gene delivery.

WO 00/39302

AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Patent Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 (published 23 January 1992) and WO 93/03769 (published 4 March 1993); Lebkowski et al., *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent et al., *Vaccines* 90 (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzychka, N. *Current Topics in Microbiol. and Immunol.* (1992) 158:97-129; Kotin, R.M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou et al., *J. Exp. Med.* (1994) 179:1867-1875.

Another vector system useful for delivering the polynucleotides of the present invention is the enterically administered recombinant poxvirus vaccines described by Small, Jr., P.A., et al. (U.S. Patent No. 5,676,950, issued October 14, 1997).

Additional viral vectors which will find use for delivering the nucleic acid molecules encoding the antigens of interest include those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the genes can be constructed as follows. The DNA encoding the particular synthetic Gag/antigen coding sequence is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the coding sequences of interest into the viral genome. The resulting TK recombinant can be selected by culturing the

cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the genes. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an avipox vector is particularly desirable in human and other mammalian species since members of the avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

Members of the Alphavirus genus, such as, but not limited to, vectors derived from the Sindbis, Semliki Forest, and Venezuelan Equine Encephalitis viruses, will also find use as viral vectors for delivering the polynucleotides of the present invention (for example, a synthetic Gag- or Env-polypeptide encoding expression cassette as described in Example 14 below). For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072; as well as, Dubensky, Jr., T.W., et al., U.S. Patent No. 5,843,723,

issued December 1, 1998, and Dubensky, Jr., T.W., U.S. Patent No. 5,789,245, issued August 4, 1998.

A vaccinia based infection/transfection system can be conveniently used to provide for inducible, transient expression of the coding sequences of interest (for example, a synthetic Gag/HCV-core expression cassette) in a host cell. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, e.g., Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

As an alternative approach to infection with vaccinia or avipox virus recombinants, or to the delivery of genes using other viral vectors, an amplification system can be used that will lead to high level expression following introduction into host cells. Specifically, a T7 RNA polymerase promoter preceding the coding region for T7 RNA polymerase can be engineered. Translation of RNA derived from this template will generate T7 RNA polymerase which in turn will transcribe more template. Concomitantly, there will be a cDNA whose expression is under the control of the T7 promoter. Thus, some of the T7 RNA polymerase generated from

translation of the amplification template RNA will lead to transcription of the desired gene. Because some T7 RNA polymerase is required to initiate the amplification, T7 RNA polymerase can be introduced into cells along with 5 the template(s) to prime the transcription reaction. The polymerase can be introduced as a protein or on a plasmid encoding the RNA polymerase. For a further discussion of T7 systems and their use for transforming cells, see, e.g., International Publication No. WO 94/26911; Studier 10 and Moffatt, *J. Mol. Biol.* (1986) 189:113-130; Deng and Wolff, *Gene* (1994) 143:245-249; Gao et al., *Biochem. Biophys. Res. Commun.* (1994) 200:1201-1206; Gao and Huang, *Nuc. Acids Res.* (1993) 21:2867-2872; Chen et al., *Nuc. Acids Res.* (1994) 22:2114-2120; and U.S. Patent No. 15 5,135,855.

The synthetic expression cassette of interest can also be delivered without a viral vector. For example, the synthetic expression cassette can be packaged as DNA or RNA in liposomes prior to delivery to the subject or 20 to cells derived therefrom. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or 25 more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta.* (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

30 Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes have been shown to mediate intracellular

delivery of plasmid DNA (Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416); mRNA (Malone et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:6077-6081); and purified transcription factors (Debs et al., *J. Biol. Chem.* (1990) 265:10189-10192), in functional form.

5 Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethyl-ammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY.

10 (See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other commercially available lipids include (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

15 Similarly, anionic and neutral liposomes are readily available, such as, from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

20 The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in METHODS OF

IMMUNOLOGY (1983), Vol. 101, pp. 512-527; Szoka et al.,
Proc. Natl. Acad. Sci. USA (1978) 75:4194-4198;
Papahadjopoulos et al., Biochim. Biophys. Acta (1975)
394:483; Wilson et al., Cell (1979) 17:77); Deamer and
5 Bangham, Biochim. Biophys. Acta (1976) 443:629; Ostro et
al., Biochem. Biophys. Res. Commun. (1977) 76:836; Fraley
et al., Proc. Natl. Acad. Sci. USA (1979) 76:3348); Enoch
and Strittmatter, Proc. Natl. Acad. Sci. USA (1979)
76:145); Fraley et al., J. Biol. Chem. (1980) 255:10431;
10 Szoka and Papahadjopoulos, Proc. Natl. Acad. Sci. USA
(1978) 75:145; and Schaefer-Ridder et al., Science (1982)
215:166.

The DNA and/or protein antigen(s) can also be delivered in cochleate lipid compositions similar to those described by Papahadjopoulos et al., Biochem. Biophys. Acta. (1975) 394:483-491. See, also, U.S. Patent Nos. 4,663,161 and 4,871,488.

The synthetic expression cassette of interest (e.g., any of the synthetic expression cassettes described in Example 1) may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected antigen to the immune system and promote migration, trapping and retention of antigens in local lymph nodes. The particles can be taken up by profession antigen presenting cells such as macrophages and dendritic cells, and/or can enhance antigen presentation through other mechanisms such as stimulation of cytokine release. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., Pharm. Res. (1993) 10:362-368; McGee JP,

et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993.

Furthermore, other particulate systems and polymers can be used for the *in vivo* or *ex vivo* delivery of the 5 gene of interest. For example, polymers such as polylysine, polyarginine, polyornithine, spermine, spermidine, as well as conjugates of these molecules, are useful for transferring a nucleic acid of interest.

Similarly, DEAE dextran-mediated transfection, calcium 10 phosphate precipitation or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like, will find use with the present methods. See, e.g., 15 Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. Peptoids (Zuckerman, R.N., et al., U.S. Patent No. 5,831,005, issued November 3, 1998) may also be used for delivery of a construct of the present 20 invention.

Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering synthetic expression cassettes of the present invention. The particles are 25 coated with the synthetic expression cassette(s) to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent 30 Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744. Also, needle-less injection systems can be used (Davis, H.L., et al.; *Vaccine* 12:1503-1509, 1994; Bioject, Inc., Portland, OR).

Recombinant vectors carrying a synthetic expression cassette of the present invention are formulated into compositions for delivery to the vertebrate subject. These compositions may either be prophylactic (to prevent infection) or therapeutic (to treat disease after infection). The compositions will comprise a "therapeutically effective amount" of the gene of interest such that an amount of the antigen can be produced *in vivo* so that an immune response is generated in the individual to which it is administered. The exact amount necessary will vary depending on the subject being treated; the age and general condition of the subject to be treated; the capacity of the subject's immune system to synthesize antibodies; the degree of protection desired; the severity of the condition being treated; the particular antigen selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. Thus, a "therapeutically effective amount" will fall in a relatively broad range that can be determined through routine trials.

The compositions will generally include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, polyethyleneglycol, hyaluronic acid, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, surfactants and the like, may be present in such vehicles. Certain facilitators of immunogenicity or of nucleic acid uptake and/or expression can also be included in the compositions or coadministered, such as, but not limited to, bupivacaine, cardiotoxin and sucrose.

Once formulated, the compositions of the invention can be administered directly to the subject (e.g., as

described above) or, alternatively, delivered *ex vivo*, to cells derived from the subject, using methods such as those described above. For example, methods for the *ex vivo* delivery and reimplantation of transformed cells 5 into a subject are known in the art and can include, e.g., dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, lipofectamine and LT-1 mediated transfection, protoplast fusion, electroporation, encapsulation of the 10 polynucleotide(s) (with or without the corresponding antigen) in liposomes, and direct microinjection of the DNA into nuclei.

Direct delivery of synthetic expression cassette compositions *in vivo* will generally be accomplished with 15 or without viral vectors, as described above, by injection using either a conventional syringe, needless devices such as Bioject® or a gene gun, such as the Accell® gene delivery system (PowderJect Technologies, Inc., Oxford, England). The constructs can be delivered 20 (e.g., injected) either subcutaneously, epidermally, intradermally, intramuscularly, intravenous, intramucosally (such as nasally, rectally and vaginally), intraperitoneally or orally. Delivery of DNA into cells of the epidermis is particularly preferred as this mode 25 of administration provides access to skin-associated lymphoid cells and provides for a transient presence of DNA in the recipient. Other modes of administration include oral ingestion and pulmonary administration, suppositories, needle-less injection, transcutaneous and 30 transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule.

2.3.2 EX VIVO DELIVERY OF THE SYNTHETIC EXPRESSION
CASSETTES OF THE PRESENT INVENTION

In one embodiment, T cells, and related cell types (including but not limited to antigen presenting cells, such as, macrophage, monocytes, lymphoid cells, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), can be used for ex vivo delivery of the synthetic expression cassettes of the present invention. T cells can be isolated from peripheral blood lymphocytes (PBLs) by a variety of procedures known to those skilled in the art. For example, T cell populations can be "enriched" from a population of PBLs through the removal of accessory and B cells. In particular, T cell enrichment can be accomplished by the elimination of non-T cells using anti-MHC class II monoclonal antibodies. Similarly, other antibodies can be used to deplete specific populations of non-T cells. For example, anti-Ig antibody molecules can be used to deplete B cells and anti-MacI antibody molecules can be used to deplete macrophages.

T cells can be further fractionated into a number of different subpopulations by techniques known to those skilled in the art. Two major subpopulations can be isolated based on their differential expression of the cell surface markers CD4 and CD8. For example, following the enrichment of T cells as described above, CD4⁺ cells can be enriched using antibodies specific for CD4 (see Coligan et al., *supra*). The antibodies may be coupled to a solid support such as magnetic beads. Conversely, CD8⁺ cells can be enriched through the use of antibodies specific for CD4 (to remove CD4⁺ cells), or can be isolated by the use of CD8 antibodies coupled to a solid support. CD4

lymphocytes from HIV-1 infected patients can be expanded *ex vivo*, before or after transduction as described by Wilson et. al. (1995) *J. Infect. Dis.* 172:88.

5 Following purification of T cells, a variety of methods of genetic modification known to those skilled in the art can be performed using non-viral or viral-based gene transfer vectors constructed as described herein. For example, one such approach involves transduction of 10 the purified T cell population with vector-containing supernatant of cultures derived from vector producing cells. A second approach involves co-cultivation of an irradiated monolayer of vector-producing cells with the purified T cells. A third approach involves a similar 15 co-cultivation approach; however, the purified T cells are pre-stimulated with various cytokines and cultured 48 hours prior to the co-cultivation with the irradiated vector producing cells. Pre-stimulation prior to such transduction increases effective gene transfer (Nolta et 20 al. (1992) *Exp. Hematol.* 20:1065). Stimulation of these cultures to proliferate also provides increased cell populations for re-infusion into the patient. Subsequent to co-cultivation, T cells are collected from the vector producing cell monolayer, expanded, and frozen in liquid 25 nitrogen.

Gene transfer vectors, containing one or more synthetic expression cassette of the present invention (associated with appropriate control elements for delivery to the isolated T cells) can be assembled using 30 known methods.

Selectable markers can also be used in the construction of gene transfer vectors. For example, a marker can be used which imparts to a mammalian cell transduced with the gene transfer vector resistance to a

cytotoxic agent. The cytotoxic agent can be, but is not limited to, neomycin, aminoglycoside, tetracycline, chloramphenicol, sulfonamide, actinomycin, netropsin, distamycin A, anthracycline, or pyrazinamide. For example, neomycin phosphotransferase II imparts resistance to the neomycin analogue geneticin (G418).

The T cells can also be maintained in a medium containing at least one type of growth factor prior to being selected. A variety of growth factors are known in the art which sustain the growth of a particular cell type. Examples of such growth factors are cytokine mitogens such as rIL-2, IL-10, IL-12, and IL-15, which promote growth and activation of lymphocytes. Certain types of cells are stimulated by other growth factors such as hormones, including human chorionic gonadotropin (hCG) and human growth hormone. The selection of an appropriate growth factor for a particular cell population is readily accomplished by one of skill in the art.

For example, white blood cells such as differentiated progenitor and stem cells are stimulated by a variety of growth factors. More particularly, IL-3, IL-4, IL-5, IL-6, IL-9, GM-CSF, M-CSF, and G-CSF, produced by activated T_h and activated macrophages, stimulate myeloid stem cells, which then differentiate into pluripotent stem cells, granulocyte-monocyte progenitors, eosinophil progenitors, basophil progenitors, megakaryocytes, and erythroid progenitors. Differentiation is modulated by growth factors such as GM-CSF, IL-3, IL-6, IL-11, and EPO.

Pluripotent stem cells then differentiate into lymphoid stem cells, bone marrow stromal cells, T cell progenitors, B cell progenitors, thymocytes, T_h Cells, T_c cells, and B cells. This differentiation is modulated by

growth factors such as IL-3, IL-4, IL-6, IL-7, GM-CSF, M-CSF, G-CSF, IL-2, and IL-5.

5 Granulocyte-monocyte progenitors differentiate to monocytes, macrophages, and neutrophils. Such differentiation is modulated by the growth factors GM-CSF, M-CSF, and IL-8. Eosinophil progenitors differentiate into eosinophils. This process is modulated by GM-CSF and IL-5.

10 The differentiation of basophil progenitors into mast cells and basophils is modulated by GM-CSF, IL-4, and IL-9. Megakaryocytes produce platelets in response to GM-CSF, EPO, and IL-6. Erythroid progenitor cells differentiate into red blood cells in response to EPO.

15 Thus, during activation by the CD3-binding agent, T cells can also be contacted with a mitogen, for example a cytokine such as IL-2. In particularly preferred embodiments, the IL-2 is added to the population of T cells at a concentration of about 50 to 100 µg/ml. Activation with the CD3-binding agent can be carried out 20 for 2 to 4 days.

Once suitably activated, the T cells are genetically modified by contacting the same with a suitable gene transfer vector under conditions that allow for transfection of the vectors into the T cells. Genetic modification is carried out when the cell density of the T cell population is between about 0.1×10^6 and 5×10^6 , preferably between about 0.5×10^6 and 2×10^6 . A number of suitable viral and nonviral-based gene transfer vectors have been described for use herein.

30 After transduction, transduced cells are selected away from non-transduced cells using known techniques. For example, if the gene transfer vector used in the transduction includes a selectable marker which confers resistance to a cytotoxic agent, the cells can be

contacted with the appropriate cytotoxic agent, whereby non-transduced cells can be negatively selected away from the transduced cells. If the selectable marker is a cell surface marker, the cells can be contacted with a binding agent specific for the particular cell surface marker, whereby the transduced cells can be positively selected away from the population. The selection step can also entail fluorescence-activated cell sorting (FACS) techniques, such as where FACS is used to select cells from the population containing a particular surface marker, or the selection step can entail the use of magnetically responsive particles as retrievable supports for target cell capture and/or background removal.

More particularly, positive selection of the transduced cells can be performed using a FACS cell sorter (e.g., a FACSVantage™ Cell Sorter, Becton Dickinson Immunocytometry Systems, San Jose, CA) to sort and collect transduced cells expressing a selectable cell surface marker. Following transduction, the cells are stained with fluorescent-labeled antibody molecules directed against the particular cell surface marker. The amount of bound antibody on each cell can be measured by passing droplets containing the cells through the cell sorter. By imparting an electromagnetic charge to droplets containing the stained cells, the transduced cells can be separated from other cells. The positively selected cells are then harvested in sterile collection vessels. These cell sorting procedures are described in detail, for example, in the FACSVantage™ Training Manual, with particular reference to sections 3-11 to 3-28 and 10-1 to 10-17.

Positive selection of the transduced cells can also be performed using magnetic separation of cells based on expression of a particular cell surface marker. In such

separation techniques, cells to be positively selected are first contacted with specific binding agent (e.g., an antibody or reagent the interacts specifically with the cell surface marker). The cells are then contacted with 5 retrievable particles (e.g., magnetically responsive particles) which are coupled with a reagent that binds the specific binding agent (that has bound to the positive cells). The cell-binding agent-particle complex can then be physically separated from non-labeled cells, 10 for example using a magnetic field. When using magnetically responsive particles, the labeled cells can be retained in a container using a magnetic field while the negative cells are removed. These and similar separation procedures are known to those of ordinary 15 skill in the art.

Expression of the vector in the selected transduced cells can be assessed by a number of assays known to those skilled in the art. For example, Western blot or Northern analysis can be employed depending on the nature 20 of the inserted nucleotide sequence of interest. Once expression has been established and the transformed T cells have been tested for the presence of the selected synthetic expression cassette, they are ready for infusion into a patient via the peripheral blood stream.

25 The invention includes a kit for genetic modification of an *ex vivo* population of primary mammalian cells. The kit typically contains a gene transfer vector coding for at least one selectable marker and at least one synthetic expression cassette contained 30 in one or more containers, ancillary reagents or hardware, and instructions for use of the kit.

EXPERIMENTAL

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

Example 1**Generation of Synthetic Gag and Env Expression Cassettes**

15 A. Modification of HIV-1 Gag, Gag-protease, Gag-reverse transcriptase and Gag-polymerase Nucleic Acid Coding Sequences

The Gag (SEQ ID NO:1), Gag-protease (SEQ ID NO:2), Gag-polymerase (SEQ ID NO:3), and Gag-reverse transcriptase (SEQ ID NO:77) coding sequences were selected from the HIV-1SF2 strain (Sanchez-Pescador, R., et al., *Science* 227(4686): 484-492, 1985; Luciw, P.A., et al. U.S. Patent No. 5,156,949, issued October 20, 1992; Luciw, P.A., et al., U.S. Patent No. 5,688,688, November 18, 1997). These sequences were manipulated to maximize expression of their gene products.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a high AU content in the RNA and in a decreased translation ability and instability of the

mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Gag-encoding sequences were modified to be comparable to codon usage found in highly expressed human genes.

5 Figure 11 presents a comparison of the percent A-T content for the cDNAs of stable versus unstable RNAs (comparison window size = 50). Human IFN γ mRNA is known to (i) be unstable, (ii) have a short half-life, and (iii) have a high A-U content. Human GAPDH (glyceraldehyde-3-phosphate dehydrogenase) mRNA is known to (i) be a stable RNA, and (i) have a low A-U content.

10 In Figure 11, the percent A-T content of these two sequences are compared to the percent A-T content of native HIV-1SF2 Gag cDNA and to the synthetic Gag cDNA sequence of the present invention. The top two panels of the figure show the percent A-T content over the length of the sequences for IFN γ and native Gag. The bottom two panels of the figure show the percent A-T content over the length of the sequences for GAPDH and the synthetic Gag.

15 Experiments performed in support of the present invention showed that the synthetic Gag sequences were capable of higher level of protein production (see the Examples) than the native Gag sequences. The data in Figure 11 suggest that one reason for this increased production may be increased stability of the mRNA corresponding to the synthetic Gag coding sequences versus the mRNA corresponding to the native Gag coding sequences.

20

25

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences (Schneider R, et al., J Virol. 71(7):4892-4903, 1997). RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating

effects of the INS. To overcome the requirement for post-transcriptional activating mechanisms of RRE and Rev, and to enhance independent expression of the Gag polypeptide, the INS were inactivated by introducing 5 multiple point mutations that did not alter the reading frame of the encoded proteins. Figure 1 shows the original SF2 Gag sequence, the location of the INS sequences, and the modifications made to the INS sequences to reduce their effects.

10 For the Gag-protease sequence (wild type, SEQ ID NO:2; synthetic, SEQ ID NOS:5, 78 and 79), the changes in codon usage were restricted to the regions up to the -1 frameshift and starting again at the end of the Gag reading frame (Figure 2; the region indicated in lower 15 case letters in Figure 2 is the unmodified region). Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence were altered as well (indicated in Figure 2). The synthetic coding sequences 20 were assembled by the Midland Certified Reagent Company (Midland, Texas).

Modification of the Gag-polymerase sequences (wild type, SEQ ID NO:3; synthetic, SEQ ID NO:6) and Gag-reverse transcriptase sequences (SEQ ID NOS:80 through 25 84) include similar modifications as described for Gag-protease in order to preserve the frameshift region. Locations of the inactivation sites and changes to the sequence to alter the inactivation sites are presented in Figure 12 for the native HIV-1_{SF2} Gag-polymerase sequence.

30 In one embodiment of the invention, the full length polymerase coding region of the Gag-polymerase sequence is included with the synthetic Gag sequences in order to increase the number of epitopes for virus-like particles expressed by the synthetic, optimized Gag expression

cassette. Because synthetic HIV-1 Gag-polymerase expresses the potentially deleterious functional enzymes reverse transcriptase (RT) and integrase (INT) (in addition to the structural proteins and protease), it is important to inactivate RT and INT functions. Several in-frame deletions in the RT and INT reading frame can be made to achieve catalytic nonfunctional enzymes with respect to their RT and INT activity. {Jay. A. Levy (Editor) (1995) *The Retroviridae*, Plenum Press, New York.

10 ISBN 0-306-45033X. Pages 215-20; Grimison, B. and Laurence, J. (1995), *Journal Of Acquired Immune Deficiency Syndromes and Human Retrovirology* 9(1):58-68; Wakefield, J. K., et al., (1992) *Journal Of Virology* 66(11):6806-6812; Esnouf, R., et al., (1995) *Nature Structural Biology* 2(4):303-308; Maignan, S., et al., (1998) *Journal Of Molecular Biology* 282(2):359-368; Katz, R. A. and Skalka, A. M. (1994) *Annual Review Of Biochemistry* 73 (1994); Jacobo-Molina, A., et al., (1993) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 90(13):6320-6324; Hickman, A. B., et al., (1994) *Journal Of Biological Chemistry* 269(46):29279-29287; Goldgur, Y., et al., (1998) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 95(16):9150-9154; Goette, M., et al., (1998) *Journal Of Biological Chemistry* 273(17):10139-10146; Gorton, J. L., et al., (1998) *Journal of Virology* 72(6):5046-5055; Engelman, A., et al., (1997) *Journal Of Virology* 71(5):3507-3514; Dyda, F., et al., *Science* 266(5193):1981-1986; Davies, J. F., et al., (1991) *Science* 252(5002):88-95; Bujacz, G., et al., (1996) *Fefs Letters* 398(2-3):175-178; Beard, W. A., et al., (1996) *Journal Of Biological Chemistry* 271(21):12213-12220; Kohlstaedt, L. A., et al., (1992)

Science 256(5065):1783-1790; Krug, M. S. and Berger, S. L. (1991) Biochemistry 30(44):10614-10623; Mazumder, A., et al., (1996) Molecular Pharmacology 49(4):621-628; Palaniappan, C., et al., (1997) Journal Of Biological Chemistry 272(17):11157-11164; Rodgers, D. W., et al., (1995) Proceedings Of the National Academy Of Sciences Of the United States Of America 92(4):1222-1226; Sheng, N. and Dennis, D. (1993) Biochemistry 32(18):4938-4942; Spence, R. A., et al., (1995) Science 267(5200):988-993.)

10 Furthermore selected B- and/or T-cell epitopes can be added to the Gag-polymerase constructs within the deletions of the RT- and INT-coding sequence to replace and augment any epitopes deleted by the functional modifications of RT and INT. Alternately, selected B-
15 and T-cell epitopes (including CTL epitopes) from RT and INT can be included in a minimal VLP formed by expression of the synthetic Gag or synthetic GagProt cassette, described above. (For descriptions of known HIV B- and T-cell epitopes see, HIV Molecular Immunology Database CTL
20 Search Interface; Los Alamos Sequence Compendia, 1987-1997; Internet address: <http://hiv-web.lanl.gov/immunology/index.html>.)

The resulting modified coding sequences are presented as a synthetic Gag expression cassette (SEQ ID NO:4), a synthetic Gag-protease expression cassette (SEQ ID NOS:5, 78 and 79), and a synthetic Gag-polymerase expression cassette (SEQ ID NO:6). Synthetic expression cassettes containing codon modifications in the reverse transcriptase region are shown in SEQ ID NOS:80 through 30 84. An alignment of selected sequences is presented in Figure 7. A common region (Gag-common; SEQ ID NO:9) extends from position 1 to position 1262.

The synthetic DNA fragments for Gag and Gag-protease were cloned into the following expression vectors:

pCMVKm2, for transient expression assays and DNA immunization studies, the pCMVKm2 vector was derived from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) and comprises a kanamycin selectable marker, a 5 ColE1 origin of replication, a CMV promoter enhancer and Intron A, followed by an insertion site for the synthetic sequences described below followed by a polyadenylation signal derived from bovine growth hormone -- the pCMVKm2 vector differs from the pCMV-link vector only in that a 10 polylinker site was inserted into pCMVKm2 to generate pCMV-link (Figure 14, polylinker at positions 1646 to 1697); pESN2dhfr (Figure 13A) and pCMVPLEdhfr (also known as pCMVIII as shown in Figure 13B), for expression in Chinese Hamster Ovary (CHO) cells; and, pAcC13, a shuttle 15 vector for use in the Baculovirus expression system (pAcC13, was derived from pAcC12 which was described by Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990).

A restriction map for vector pCMV-link is presented 20 in Figure 14. In the figure, the CMV promoter (CMV IE ENH/PRO), bovine growth hormone terminator (BGH pA), kanamycin selectable marker (kan), and a ColE1 origin of replication (ColE1 ori) are indicated. A polycloning site is also indicated in the figure following the CMV 25 promoter sequences.

A restriction map for vector pESN2dhfr is presented in Figure 13A. In the figure, the CMV promoter (pCMV, hCMVIE), bovine growth hormone terminator (BGH_pA), SV40 30 origin of replication (SV40ori), neomycin selectable marker (Neo), SV40 polyA (SV40pA), Adenovirus 2 late promoter (Ad2VLP), and the murine dhfr gene (μ dhfr) are indicated. A polycloning site is also indicated in the figure following the CMV promoter sequences.

Briefly, construction of pCMVPLEdhfr (pCMVIII) was as follows. To construct a DHFR cassette, the EMCV IRES (internal ribosome entry site) leader was PCR-amplified from pCite-4a+ (Novagen, Inc., Milwaukee, WI) and inserted into pET-23d (Novagen, Inc., Milwaukee, WI) as an Xba-Nco fragment to give pET-EMCV. The dhfr gene was PCR-amplified from pESN2dhfr to give a product with a Gly-Gly-Gly-Ser spacer in place of the translation stop codon and inserted as an Nco-BamH1 fragment to give pET-E-DHFR. Next, the attenuated neo gene was PCR amplified from a pSV2Neo (Clontech, Palo Alto, CA) derivative and inserted into the unique BamH1 site of pET-E-DHFR to give pET-E-DHFR/Neo_(m2). Then, the bovine growth hormone terminator from pCDNA3 (Invitrogen, Inc., Carlsbad, CA) was inserted downstream of the neo gene to give pET-E-DHFR/Neo_(m2)BGHt. The EMCV-dhfr/neo selectable marker cassette fragment was prepared by cleavage of pET-E-DHFR/Neo_(m2)BGHt. The CMV enhancer/promoter plus Intron A was transferred from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) as a HindIII-SalI fragment into pUC19 (New England Biolabs, Inc., Beverly, MA). The vector backbone of pUC19 was deleted from the NdeI to the Sapi sites. The above described DHFR cassette was added to the construct such that the EMCV IRES followed the CMV promoter to produce the final construct. The vector also contained an amp^r gene and an SV40 origin of replication.

Selected pCMVKm2 vectors containing the synthetic expression cassettes have been designated as follows: pCMVKm2.GagMod.SF2, pCMVKm2.GagprotMod.SF2, and pCMVKm2.GagpolMod.SF2, pCMVKm2.GagprotMod.SF2.GP1 (SEQ ID NO:78) and pCMVKm2.GagprotMod.SF2.GP2 (SEQ ID NO:79). Other exemplary Gag-encoding expressing cassettes are shown in the Figures and as Sequence Listings.

B. Modification of HIV-1 Gag/Hepatitis C Core Chimeric Protein Nucleic Acid Coding Sequences Generation of Synthetic Expression Cassettes

To facilitate the ligation of the Gag and HCV core coding sequences, PCR amplification was employed. The synthetic p55Gag expression cassette was used as a PCR template with the following primers: GAG5 (SEQ ID NO:11) and P55-SAL3 (SEQ ID NO:12). The PCR amplification was conducted at 55°C for 25 cycles using Stratagene's Pfu polymerase. The resulting PCR product was rendered free of nucleotides and primers using the Promega PCR clean-up kit and then subjected to EcoRI and SalI digestions. For HCV core coding sequences, the following primers were used with an HCV template (Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997): CORESAL 5 (SEQ ID NO:13) and 173CORE (SEQ ID NO:14) using the conditions outlined above. The purified product was digested with SalI and BamHI restriction enzymes. The digested Gag and HCV core PCR products were ligated into the pCMVKm2 vector digested with EcoRI and BamHI. Ligation of the PCR products at the SalI site resulted in a direct fusion of the final amino acid of p55Gag to the second amino acid of HCV core, serine. Amino acid 173 of core is a serine and is followed immediately by a TAG termination codon. The sequence of the fusion clone was confirmed. The pCMVKm2 vector containing the synthetic expression

cassette was designated as pCMVKm2.GagModHCVcore.

The EcoRI-BamHI fragment of p55Gag-core 173 was also cloned into EcoRI-BamHI-digested pAcC13 for baculovirus expression. Western blots confirmed expression and sucrose gradient sedimentation along with electron microscopy confirmed particle formation. To generate the above clone but containing the synthetic Gag sequences (instead of wild-type), the following steps were performed: pCMVKm2-modified p55Gag was used as template for PCR amplification with MS65 (SEQ ID NO:15) and MS66 (SEQ ID NO:16) primers. The region amplified corresponds to the BspHI and SalI sites at the C-terminus of synthetic Gag sequence. The amplification product was digested with BspHI and SalI and ligated to SalI/BamHI digested pCMV-link along with the Sal/BspHI fragment from pCMV-Km-p55modGag , representing the amino terminal end of modified Gag, and the SalI/BamHI fragment from pCMV-p55Gag-core173. Thereafter, a T4-blunted-SalI partial/BamHI fragment was ligated into pAcC4-SmaI/BamHI to generate pAcC4-p55GagMod-core173 (containing the synthetic sequence presented as SEQ ID NO:7).

C. Defining of the Major Homology Region (MHR) of HIV-1 p55Gag

The Major Homology Region (MHR) of HIV-1 p55 (Gag) is located in the p24-CA sequence of Gag. It is a conserved stretch of 20 amino acids (SEQ ID NO:19). The position in the wild type HIV-1_{sf2} Gag protein is from aa 286-305 and spans a region from nucleotides 856-915 in the native HIV-1_{sf2} Gag DNA-sequence. The position in the synthetic Gag protein is from aa 288-307 and spans a region from nucleotides 862-921 for the synthetic Gag DNA-sequence. The nucleotide sequence for the MHR in the synthetic

GagMod.SF2 is presented as SEQ ID NO:20. Mutations or deletions in the amino acid sequence of the MHR can severely impair particle production (Borsetti, A., et al., J. Virol. 72(11):9313-9317, 1998; Mammano, F., et al., J Virol 68(8):4927-4936, 1994).

Percent identity to the MHR nucleotide sequence can be determined, for example, using the MacDNAsis program (Hitachi Software Engineering America Limited, South San Francisco, CA), Higgins algorithm, with the following exemplary parameters: gap penalty = 5, no. of top diagonals = 5, fixed gap penalty = 5, K-tuple = 2, window size = 5, and floating gap penalty = 10.

D. Generation of Synthetic Env Expression Cassettes

Env coding sequences of the present invention include, but are not limited to, polynucleotide sequences encoding the following HIV-encoded polypeptides: gp160, gp140, and gp120 (see, e.g., U.S. Patent No. 5,792,459 for a description of the HIV-1_{SF2} ("SF2") Env polypeptide). The relationships between these polypeptides is shown schematically in Figure 15 (in the figure: the polypeptides are indicated as lines, the amino and carboxy termini are indicated on the gp160 line; the open circle represents the oligomerization domain; the open square represents a transmembrane spanning domain (TM); and "c" represents the location of a cleavage site, in gp140 mut the "X" indicates that the cleavage site has been mutated such that it no longer functions as a cleavage site). The polypeptide gp160 includes the coding sequences for gp120 and gp41. The polypeptide gp41 is comprised of several domains including an oligomerization domain (OD) and a transmembrane spanning domain (TM). In the native envelope, the oligomerization domain is required for the

non-covalent association of three gp41 polypeptides to form a trimeric structure: through non-covalent interactions with the gp41 trimer (and itself), the gp120 polypeptides are also organized in a trimeric structure.

5 A cleavage site (or cleavage sites) exists approximately between the polypeptide sequences for gp120 and the polypeptide sequences corresponding to gp41. This cleavage site(s) can be mutated to prevent cleavage at the site. The resulting gp140 polypeptide corresponds to

10 a truncated form of gp160 where the transmembrane spanning domain of gp41 has been deleted. This gp140 polypeptide can exist in both monomeric and oligomeric (i.e. trimeric) forms by virtue of the presence of the oligomerization domain in the gp41 moiety. In the

15 situation where the cleavage site has been mutated to prevent cleavage and the transmembrane portion of gp41 has been deleted the resulting polypeptide product is designated "mutated" gp140 (e.g., gp140.mut). As will be apparent to those in the field, the cleavage site can be

20 mutated in a variety of ways. The native amino acid sequence in the SF162 cleavage sites is: APTKAKRRVVQREKR (SEQ ID NO:21), where KAKRR (SEQ ID NO:22) is termed the "second" site and REKR (SEQ ID NO:23) is the "first site". Exemplary mutations include the following

25 constructs: gp140.mut7.modSF162 which encodes the amino acid sequence APTKA**I**SSVVQSEKS (SEQ ID NO:24) in the cleavage site region; gp140.mut8.modSF162 which encodes the amino acid sequence APTIA**I**SSVVQSEKS (SEQ ID NO:25) in the cleavage site region and gp140mut.modSF162 which

30 encodes the amino acid sequence APTKAKRRVVQREKS (SEQ ID NO:26). Mutations are denoted in bold. The native amino acid sequence in the US4 cleavage sites is:
APT**Q**A**K**RRVVQREKR (SEQ ID NO:27), where QAKRR (SEQ ID NO:28) is termed the "second" site and REKR (SEQ ID

NO:23) is the "first site". Exemplary mutations include the following construct: gp140.mut.modUS4 which encodes the amino acid sequence APTQAKRRVVQREKS (SEQ ID NO:29) in the cleavage site region. Mutations are denoted in bold.

5

E. Modification of HIV-1 Env (Envelope) Nucleic Acid Coding Sequences

In one embodiment of the present invention, wild-type Env coding sequences were selected from the HIV-1_{SF162} ("SF162") strain (Cheng-Mayer (1989) PNAS USA 86:8575-8579). These SF162 sequences were as follows: gp120, SEQ ID NO:30 (Fig. 16); gp140, SEQ ID NO:31 (Fig. 17); and gp160, SEQ ID NO:32 (Fig. 18).

In another embodiment of the present invention, wild-type Env coding sequences were selected from the HIV-US4 strain (Mascola, et al. (1994) J. Infect. Dis. 169:48-54). These US4 sequences were as follows: gp120, SEQ ID NO:51 (Fig. 38); gp140, SEQ ID NO:52 (Fig. 39); and gp160, SEQ ID NO:53 (Fig. 40).

These Env coding sequences were manipulated to maximize expression of their gene products.

First, the wild-type coding region was modified in one or more of the following ways. In one embodiment, sequences encoding hypervariable regions of Env, particularly V1 and/or V2 were deleted. In other embodiments, mutations were introduced into sequences encoding the cleavage site in Env to abrogate the enzymatic cleavage of oligomeric gp140 into gp120 monomers. (See, e.g., Earl et al. (1990) PNAS USA 87:648-652; Earl et al. (1991) J. Virol. 65:31-41). In yet other embodiments, hypervariable region(s) were deleted, N-glycosylation sites were removed and/or cleavage sites mutated.

Second, the HIV-1 codon usage pattern was modified

so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T in the codon-triplet. The effect of 5 the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Env coding sequences were modified to be comparable 10 to codon usage found in highly expressed human genes.

Figures 22A-22H present comparisons of the percent A-T content for the cDNAs of stable versus unstable RNAs (comparison window size = 50). Human IFN γ mRNA is known to (i) be unstable, (ii) have a short half-life, and 15 (iii) have a high A-U content. Human GAPDH (glyceraldehyde-3-phosphate dehydrogenase) mRNA is known to (i) be a stable RNA, and (i) have a low A-U content. In Figures 22A-H, the percent A-T content of these two sequences are compared to the percent A-T content of (1) 20 native HIV-1 US4 Env gp160 cDNA, a synthetic US4 Env gp160 cDNA sequence (i.e., having modified codons) of the present invention; and (2) native HIV-1 SF162 Env gp160 cDNA, a synthetic SF162 Env gp160 cDNA sequence (i.e., having modified codons) of the present invention. 25 Figures 22A-H show the percent A-T content over the length of the sequences for IFN γ (Figures 22C and 22G); native gp160 Env US4 and SF162 (Figures 22A and 22E, respectively); GAPDH (Figures 22D and 22H); and the synthetic gp160 Env for US4 and SF162 (Figures 22B and 22F). Experiments performed in support of the present 30 invention showed that the synthetic Env sequences were capable of higher level of protein production (see the Examples) than the native Env sequences. The data in Figures 22A-H suggest that one reason for this increased

production is increased stability of the mRNA corresponding to the synthetic Env coding sequences versus the mRNA corresponding to the native Env coding sequences.

5 To create the synthetic coding sequences of the present invention the gene cassettes were designed to comprise the entire coding sequence of interest. Synthetic gene cassettes were constructed by oligonucleotide synthesis and PCR amplification to 10 generate gene fragments. Primers were chosen to provide convenient restriction sites for subcloning. The resulting fragments were then ligated to create the entire desired sequence which was then cloned into an appropriate vector. The final synthetic sequences were 15 (i) screened by restriction endonuclease digestion and analysis, (ii) subjected to DNA sequencing in order to confirm that the desired sequence had been obtained and (iii) the identity and integrity of the expressed protein confirmed by SDS-PAGE and Western blotting (See, 20 Examples. The synthetic coding sequences were assembled at Chiron Corp. or by the Midland Certified Reagent Company (Midland, Texas).

Exemplary modified coding sequences are presented as synthetic Env expression cassettes in Table 1A and 1B.

25 The following expression cassettes (i) have unique, terminal EcoRI and XbaI cloning sites; (ii) include Kozak sequences to promote optimal translation; (iii) tPA signal sequences (to direct the ENV polypeptide to the cell membrane, see, e.g., Chapman et al., *infra*); (iv) 30 open reading frames optimized for expression in mammalian cells; and (v) a translational stop signal codon.

Table 1A: Exemplary Synthetic Env Expression
Cassettes (SF162)

	Expression Cassette	Seq Id	Further Information
5	gp120 SF162	30	wild-type; Figure 16
	gp140 SF162	31	wild-type; Figure 17
	gp160 SF162	32	wild-type; Figure 18
	gp120.modSF162	33	none; Figure 19
	gp120.modSF162.delV2	34	deleted V2 loop; Figure 20
10	gp120.modSF162.delV1/V2	35	deleted V1 and V2; Figure 21
	gp140.modSF162	36	none; Figure 23
	gp140.modSF162.delV2	37	deleted V2 loop; Figure 24
	gp140.modSF162.delV1/V2	38	deleted V1 and V2; Figure 25
	gp140.mut.modSF162	39	mutated cleavage site; Fig. 26
15	gp140.mut.modSF162.delV2	40	deleted V2; mutated cleavage site; Figure 27
	gp140.mut.modSF162.delV1/V2	41	deleted V1 & V2; mutated cleavage site; Figure 28
	gp140.mut7.modSF162	42	mutated cleavage site; Fig. 29
	gp140.mut7.modSF162.delV2	43	mutated cleavage site; deleted V2; Figure 30
20	gp140.mut7.modSF162.delV1/V2	44	mutated cleavage site; deleted V1 and V2; Figure 31
	gp140.mut8.modSF162	45	mutated cleavage site; Fig. 32
	gp140.mut8.modSF162.delV2	46	mutated cleavage site; deleted V2; Figure 33
	gp140.mut8.modSF162.delV1/V2	47	mutated cleavage site; deleted V1 and V2; Figure 34
25	gp160.modSF162	48	none; Figure 35
	gp160.modSF162.delV2	49	deleted V2 loop; Figure 36
	gp160.modSF162.delV1/V2	50	deleted V1 & V2; Figure 37

Table 1B:
Exemplary Synthetic Env Expression Cassettes (US4)

	Expression Cassette	Seq Id	Further Information
5	gp120 US4	51	wild-type; Figure 38
	gp140 US4	52	wild-type; Figure 39
	gp160 US4	53	wild-type; Figure 40
	gp120.modUS4	54	none; Figure 41
	gp120.modUS4.del 128-194	55	deletion in V1 and V2 regions; Figure 42
10	gp140.modUS4	56	none; Figure 43
	gp140.mut.modUS4	57	mutated cleavage site; Figure 44
	gp140TM.modUS4	58	native transmembrane region; Figure 45
	gp140.modUS4.delV1/V2	59	deleted V1 and V2; Figure 46
	gp140.modUS4.delV2	60	deleted V1; Figure 47
	gp140.mut.modUS4.delV1/V2	61	mutated cleavage site; deleted V1 and V2; Figure 48
15	gp140.modUS4.del 128-194	62	deletion in V1 and V2 regions; Figure 49
	gp140.mut.modUS4.del 128-194	63	mutated cleavage site; deletion in V1 and V2 regions; Figure 50
	gp160.modUS4	64	none; Figure 51
	gp160.modUS4.delV1	65	deleted V1; Figure 52
20	gp160.modUS4.delV2	66	deleted V2; Figure 53
	gp160.modUS4.delV1/V2	67	deleted V1 and V2; Figure 54
	gp160.modUS4del 128-194	68	deletion in V1 and V2 regions; Figure 55

Alignments of the sequences presented in the above
25 tables are presented in Figures 66A and 66B.

A common region (Env-common) extends from nucleotide position 1186 to nucleotide position 1329 (SEQ ID NO:69,

Fig. 56) relative to the wild-type US4 sequence and from nucleotide position 1117 to position 1260 (SEQ ID NO:79, Fig. 57) relative to the wild-type SF162 sequence. The synthetic sequences of the present invention
5 corresponding to these regions are presented, as SEQ ID NO:71 (Figure 58) for the synthetic Env US4 common region and as SEQ ID NO:72 (Figure 59) for the synthetic Env SF162 common region.

Percent identity to this sequence can be determined,
10 for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5, reporting threshold = 1; alignment threshold = 20.

15 Various forms of the different embodiments of the present invention (e.g., constructs) may be combined.

F. Cloning Synthetic Env Expression Cassettes of the Present Invention.

20 The synthetic DNA fragments encoding the Env polypeptides were typically cloned into the eucaryotic expression vectors described above for Gag, for example, pCMVKm2/pCMVlink (Figure 4), pCMV6a, pESN2dhfr (Figure 13A), pCMVIII (Figure 13B; alternately designated as the
25 pCMV-PL-E-dhfr/neo vector).

Exemplary designations for pCMVlink vectors containing synthetic expression cassettes of the present invention are as follows: pCMVlink.gp140.modsSF162;
30 pCMVlink.gp140.-modsSF162.delV2;
pCMVlink.gp140.mut.modsSF162;
pCMVlink.gp140.mut.modsSF162.delV2; pCMVKm2.gp140.modUS4;
pCMVKm2.gp140.modUS4.delV2; pCMVKm2.gp140.mut.modUS4;
and, pCMVKm2.gp140.mut.modUS4.delV1/V2.

G. Generation of Synthetic Tat Expression Cassettes

Tat coding sequences have also been modified according to the teachings of the present specification. The wild type nucleotide sequence encoding tat from 5 variant SF162 is presented in Figure 76 (SEQ ID NO:85). The corresponding wild-type amino acid sequence is presented in Figure 77 (SEQ ID NO:86). Figure 81 (SEQ ID NO:89) shows the nucleotide sequence encoding the amino terminal of the tat protein and the codon encoding 10 cysteine-22 is underlined. Other exemplary constructs encoding synthetic tat polypeptides are shown in Figures 78 and 79 (SEQ ID NOs:87 and 88). In one embodiment (SEQ ID NO:88), the cystein residue at position 22 is replaced by a glycine. Caputo et al. (1996) Gene Therapy 3:235. 15 have shown that this mutation affects the trans activation domain of Tat.

Various forms of the different embodiments of the invention, described herein, may be combined.

H. Deposit of Vectors

Selected exemplary constructs shown below and described herein are deposited at Chiron Corporation, Emeryville, CA, 94662-8097, and were sent to the American Type Culture Collection, 10801 University Boulevard, 25 Manassas, VA 20110-2209 on December 27, 1999.

	Plasmid Name	Chiron	Date Sent
		Deposit #	to ATCC
	pCMVgp160.modUS4	5094	27 Dec 99
	pCMVgp160delI.modUS4	5095	27 Dec 99
	pCMVgp160del2.modUS4	5096	27 Dec 99
5	pCMVgp160del-2.modUS4	5097	27 Dec 99
	pCMVgp160del128-194.mod.US4	5098	27 Dec 99
	pCMVgp140mut.modUS4del128-194	5100	27 Dec 99
	pCMVgp140.mut.mod.US	5101	27 Dec 99
	pCMVgp160.modSF162	5125	27 Dec 99
10	pCMVgp160.modSF162.delV2	5126	27 Dec 99
	pCMVgp160.modSF162.delV1V2	5127	27 Dec 99
	pCMVgp140.mut.modSF162delV2	5128	27 Dec 99
	pCMVgp140.mut7.modSF162	5129	27 Dec 99
	pCMVgp140.mut7.modSF162delV2	5130	27 Dec 99
15	pCMVgp140.mut8.modSF162	5131	27 Dec 99
	pCMVgp140.mut8.modSF162delV2	5132	27 Dec 99
	pCMVgp140.mut8.modSF162delV1V2	5133	27 Dec 99
	pCMVKm2.Gagprot.Mod.SF2.GP1	5150	27 Dec 99
	pCMVKm2.Gagprot.Mod.SF2.GP2	5151	27 Dec 99
20			

Example 2Expression Assays for theSynthetic Gag, Env and Tat Coding Sequences25 A. Gag and Gag-Protease Coding Sequences

The HIV-1SF2 wild-type Gag (SEQ ID NO:1) and Gag-protease (SEQ ID NO:2) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Gag (SEQ ID NO:4) and Gag-protease (SEQ ID NOs:5, 78 or 79)) sequences were cloned.

Expression efficiencies for various vectors carrying the HIV-1SF2 wild-type and synthetic Gag sequences were evaluated as follows. Cells from several mammalian cell lines (293, RD, COS-7, and CHO; all obtained from the 5 American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209) were transfected with 2 µg of DNA in transfection reagent LT1 (PanVera Corporation, 545 Science Dr., Madison, WI). The cells were incubated for 5 hours in reduced serum medium (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium was then 10 replaced with normal medium as follows: 293 cells, IMDM, 10% fetal calf serum, 2% glutamine (BioWhittaker, Walkersville, MD); RD and COS-7 cells, D-MEM, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, 15 Gaithersburg, MD); and CHO cells, Ham's F-12, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The cells were incubated for either 48 or 60 hours. Supernatants were harvested and filtered 20 through 0.45 µm syringe filters and, optionally, stored at -20°C.

Supernatants were evaluated using the Coulter p24-assay (Coulter Corporation, Hialeah, FL, US), using 96-well plates coated with a murine monoclonal antibody directed against HIV core antigen. The HIV-1 p24 antigen binds to the coated wells. Biotinylated antibodies 25 against HIV recognize the bound p24 antigen. Conjugated strepavidin-horseradish peroxidase reacts with the biotin. Color develops from the reaction of peroxidase with TMB substrate. The reaction is terminated by 30 addition of 4N H₂SO₄. The intensity of the color is directly proportional to the amount of HIV p24 antigen in a sample.

The results of these expression assays are presented in Tables 2A and 2B. Tables 2A and 2B shows data

obtained using the synthetic Gag-protease expression cassette of SEQ ID NO:5. Similar results were obtained using the Gag-protease expression cassettes of SEQ ID NOS:78 and 79.

Table 2: in vitro gag and gagprot p24 expression

5 TABLE 2a. Increased in vitro expression from modified vs. native gag plasmids in supernatants and lysates from transiently transfected cells

experiment	native (nat) ^a modified (mod) ^b	supernatant (sup) lysate (lys)	cell line	hours post transfection	total ng p24 (fold increase)
1	nat	sup	293	48	3.4
	mod	sup	293	48	1260 (371)
	nat	sup	293	60	3.2
	mod	sup	293	60	2222 (694)
2	nat	sup	293	60	1.8
	mod	sup	293	60	1740 (966)
3	nat	sup	293	60	1.8
	mod	sup	293	60	580 (322)
4	nat	lys	293	60	1.5
	mod	lys	293	60	85 (57)
1	nat	sup	RD	48	5.6
	mod	sup	RD	48	66 (12)
	nat	sup	RD	60	7.8
	mod	sup	RD	60	70.2 (9)
2	nat	lys	RD	60	1.9
	mod	lys	RD	60	7.8 (4)
1	nat	sup	COS-7	48	0.4
	mod	sup	COS-7	48	33.4 (84)
2	nat	sup	COS-7	48	0.4
	mod	sup	COS-7	48	10 (25)
	nat	lys	COS-7	48	3
	mod	lys	COS-7	48	14 (5)

^apCMVLink.Gag.SF2.PRE

^bpCMVKm2.GagMod.SF2

TABLE 2b. *In vitro* expression from modified gag and gagprotease plasmids in supernatants and lysates from transiently transfected cells

5

plasmid	supernatant (sup) lysate (lys)	cell line	hours post transfection	total ng p24 ^d
Gag ^a	sup	293	60	760
GagProt(GP1) ^b	sup	293	60	380
GagProt(GP2) ^c	sup	293	60	320
Gag	lys	293	60	78
GagProt(GP1)	lys	293	60	1250
GagProt(GP2)	lys	293	60	400
Gag	sup	COS-7	72	40
GagProt(GP1)	sup	COS-7	72	150
GagProt(GP2)	sup	COS-7	72	290
Gag	lys	COS-7	72	60
GagProt(GP1)	lys	COS-7	72	63
GagProt(GP2)	lys	COS-7	72	58

^a pCMVKm2.GagMod.SF2^b pCMVKm2.GagProtMod.SF2(GP1) gagprotease with codon optimization and inactivation of INS in protease^c pCMVKm2.GagProtMod.SF2(GP2) gagprotease with only inactivation of INS in protease^d Shown are representative results from 3 independent experiments for each cell line tested.

The data showed that the synthetic Gag and Gag-protease expression cassettes provided dramatic increases in production of their protein products, relative to the native (HIV-1SF2 wild-type) sequences, when expressed in 5 a variety of cell lines.

B. Env Coding Sequences

The HIV-SF162 ("SF162") wild-type Env (SEQ ID NO:1-3) and HIV-US4 ("US4") wild-type Env (SEQ ID NO:22-24) 10 sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Env sequences were cloned.

Expression efficiencies for various vectors carrying the SF162 and US4 wild-type and synthetic Env sequences 15 were evaluated essentially as described above for Gag except that cell lysates were prepared in 40 μ l lysis buffer (1.0 % NP40, 0.1 M Tris pH 7.5) and frozen at -20°C and capture ELISAs were performed as follows.

For Capture ELISAs, 250 ng of an ammonium sulfate 20 IgG cut of goat polyclonal antibody to gp120SF2/env2-3 was used to coat each well of a 96-well plate (Corning, Corning, NY). Serial dilutions of gp120/SF2 protein (MID 167) were used to set the quantitation curve from which expression of US4 or SF162 gp120 proteins from 25 transfection supernatant and lysates were calculated.

Samples were screened undiluted and, optionally, by serial 2-fold dilutions. A human polyclonal antibody to HIV-1 gp120/SF2 was used to detect bound gp120 envelope 30 protein, followed by horse-radish peroxidase (HRP)-labeled goat anti-human IgG conjugates. TMB (Pierce, Rockford, IL) was used as the substrate and the reaction is terminated by addition of 4N H₂SO₄. The reaction was quantified by measuring the optical density (OD) at 450 nm. The intensity of the color is directly

proportional to the amount of HIV gp120 antigen in a sample. Purified SF2 gp120 protein was diluted and used as a standard.

The results of the transient expression assays are
5 presented in Tables 3 and 4. Table 3 depicts transient expression in 293 cells transfected with a pCMVKm2 vector carrying the Env cassette of interest. Table 4 depicts transient expression in RD cells transfected with a pCMVKm2 vector carrying the Env cassette of interest.

5

Table 3

Native (N) Synthetic (S)	Cell Line	Total sup (ng)	Sup fold increase (S v. N)	Total cell lysate (ng)	Cell lysate fold increase (S v. N)	Total (ng)	Total fold increase (S v. N)
N-gp120.US4	RD	87	<1			88	
S-gp120.modus4	RD	690	8	2	5	693	8
N-gp140.US4	RD	526	0			526	
S-gp140.modus4	RD	1305	2	1	2	1306	2
S-gp140mt.modus4	RD	35	N/A	25	N/A	60	N/A
S-gp140TM.modus4	RD	0	N/A	5	N/A	5	N/A
N-gp160.US4	RD	0		8		8	
S-gp160.modus4	RD	0	0	30	4	30	4

Table 4.

CHO Cell Lines Expression Level of US4 Envelope Constructs				
	Constructs	CHO Clone #	MTX Level	Expression Level (ng/ml)
5	gp120.modUS4	1	3.2μM	250-450
		2	1.6μM	350-450
		3	200nM	230-580
		4	200nM	300-500
10	gp140.modUS4	1	1μM	155-300
		2	1μM	100-260
		3	1μM	200-430
15	gp140.mut. modUS4	1	1μM	110-270
		2	1μM	100-235
		3	1μM	100-220
	gp140.modUS4 .delV1/V2	1	50nM	313-587**
		2	50nM	237-667**
		3	50nM	492-527**
	gp140.mut. modUS4.delV1 /V2	1	50nM	46-328**
		2	50nM	82-318**
		3	50nM	204-385**

*All samples measured at T-75 flask stage unless otherwise indicated

**at 24 well and 6 well plate stages

***in a three liter bioreactor perfusion culture this clone yielded approximately 2-5 μg/ml.

The data showed that the synthetic Env and expression cassettes provided a significant increase in production of their protein products, relative to the native (HIV-1SF162 or US4 wild-type) sequences, when
5 expressed in a variety of cell lines.

C. CHO Cell line Env expression data

Chinese hamster ovary (CHO) cells were transfected with plasmid DNA encoding the synthetic HIV-1 gp120 or
10 gp140 proteins (e.g., pESN2dhfr or pCMVIII vector backbone) using Mirus TransIT-LT1 polyamine transfection reagent (Pan Vera) according to the manufacturers instructions and incubated for 96 hours. After 96 hours, media was changed to selective media (F12 special with
15 250 µg/ml G418) and cells were split 1:5 and incubated for an additional 48 hours. Media was changed every 5-7 days until colonies started forming at which time the colonies were picked, plated into 96 well plates and screened by gp120 Capture ELISA. Positive clones were
20 expanded in 24 well plates and screened several times for Env protein production by Capture ELISA, as described above. After reaching confluence in 24 well plates, positive clones were expanded to T25 flasks (Corning, Corning, NY). These were screened several times after
25 confluence and positive clones were expanded to T75 flasks.

Positive T75 clones were frozen in LN2 and the highest expressing clones amplified with 0-5 µM methotrexate (MTX) at several concentrations and plated in
30 100mm culture dishes. Plates were screened for colony formation and all positive clones were again expanded as described above. Clones were expanded and amplified and screened at each step by gp120 capture ELISA. Positive clones were frozen at each methotrexate level. Highest

producing clones were grown in perfusion bioreactors (3L, 100L) for expansion and adaptation to low serum suspension culture conditions for scale-up to larger bioreactors.

5 Tables 5 and 6 show Capture ELISA data from CHO cells transfected with pCMVIII vector carrying a cassette encoding synthetic HIV-US4 and SF162 Env polypeptides (e.g., mutated cleavage sites, modified codon usage and/or deleted hypervariable regions). Thus, stably 10 transfected CHO cell lines which express Env polypeptides (e.g., gp120, gp140-monomeric, and gp140-oligomeric) have been produced.

Table 5

CHO Cell Lines Expression Level of US4 Envelope Constructs				
	Constructs	CHO Clone #	MTX Level	Expression Level (ng/ml)
5	gp120.modUS4	1	3.2 μM	250-450
		2	1.6 μM	350-450
		3	200nM	230-580***
		4	200nM	300-500
10	gp140.modUS4	1	1 μM	155-300
		2	1 μM	100-260
		3	1 μM	200-430
15	gp140.mut. modUS4	1	1 μM	110-270
		2	1 μM	100-235
		3	1 μM	100-220
	gp140.modUS4 .delV1/V2	1	50nM	313-587**
		2	50nM	237-667**
		3	50nM	492-527**
	gp140.mut. modUS4.delV1 /V2	1	50nM	46-328**
		2	50nM	82-318**
		3	50nM	204-385**

All samples measured at T-75 flask stage unless otherwise indicated

**at 24 well and 6 well plate stages

***in a three liter bioreactor perfusion culture this clone yielded approximately 2-5 μg/ml.

Table 6

CHO Cell Lines Expression Level of SF162 Envelope Constructs				
	Constructs	CHO Clone #	MTX Level	Expression Level (ng/ml)
5	gp120.modSF162	1	0	755-2705
		2	0	928-1538
		3	0	538-1609
	gp140.modSF162	1	20 nM	180-350
	gp140.mut. modSF162	1	20 nM	164-451
		2	20 nM	188-487
		3	20 nM	233-804
	gp120.modSF162 .delV2	1	800nM	528-1560
		2	800nM	487-1878
		3	800nM	589-1212
10	gp140.modSF162 .delV2	1	800nM	300-600
		2	800nM	200-400
		3	800nM	200-500
	gp140.mut. modSF162.delV2	1	800nM	300-700
		2	400nM	1161
		3	800nM	400-600
		4	400nM	1600-2176

15 *All samples measured at T-75 flask stage unless otherwise indicated

20 The results presented above demonstrate the ability of the constructs of the present invention to provide expression of Env polypeptides in CHO cells. Production of polypeptides using CHO cells provides (i) correct glycosylation patterns and protein conformation (as determined by binding to panel of MAbs); (ii) correct binding to CD4 receptor molecules; (iii) absence of non-

mammalian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification.

D. Tat Coding Sequences

5 The HIV-SF162 ("SF162") wild-type Tat (SEQ ID NO:85) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Tat sequences were cloned (SEQ ID NOS:87, 88 and 89).

10 Expression efficiencies for various vectors carrying the SF162 wild-type and synthetic Tat sequences are evaluated essentially as described above for Gag and Env using capture ELISAs with the appropriate anti-tat antibodies and/or CHO cell assays. Expression of the polypeptides encoded by the synthetic cassettes is
15 improved relative to wild type.

Example 3

Western Blot Analysis of Expression

A. Gag and Gag-Protease Coding Sequences

20 Human 293 cells were transfected as described in Example 2 with pCMV6a-based vectors containing native or synthetic Gag expression cassettes. Cells were cultivated for 60 hours post-transfection. Supernatants were prepared as described. Cell lysates were prepared
25 as follows. The cells were washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO) in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego, CA) were loaded with 20 μ l of supernatant or 12.5 μ l of cell lysate. A protein standard was also loaded (5 μ l, broad size range standard; BioRad Laboratories, Hercules, CA). Electrophoresis was carried out and the proteins were transferred using a BioRad Transfer Chamber (BioRad

Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer (Millipore), where the transfer was performed at 100 volts for 90 minutes. The 5 membranes were exposed to HIV-1-positive human patient serum and immunostained using o-phenylenediamine dihydrochloride (OPD; Sigma).

The results of the immunoblotting analysis showed that cells containing the synthetic Gag expression 10 cassette produced the expected p55 protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag p55 protein was seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants 15 for cells transfected with the synthetic Gag expression cassette of the present invention. Experiments performed in support of the present invention suggest that cells containing the synthetic Gag-prot expression cassette produced the expected Gag-prot protein at comparably 20 higher per-cell concentrations than cells containing the native expression cassette.

In addition, supernatants from the transfected 293 cells were fractionated on sucrose gradients. Aliquots of the supernatant were transferred to Polyclear™ ultracentrifuge tubes (Beckman Instruments, Columbia, MD), under-laid with a solution of 20% (wt/wt) sucrose, and 25 subjected to 2 hours centrifugation at 28,000 rpm in a Beckman SW28 rotor. The resulting pellet was suspended in PBS and layered onto a 20-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in 30 a Beckman SW41ti rotor.

The gradient was then fractionated into approximately 10 x 1 ml aliquots (starting at the top, 20%-end, of the gradient). Samples were taken from

fractions 1-9 and were electrophoresed on 8-16% SDS polyacrylamide gels. Fraction number 4 (the peak fraction) corresponds to the expected density of Gag protein VLPs. The supernatants from 293/synthetic Gag cells gave much stronger p55 bands than supernatants from 293/native Gag cells, and, as expected, the highest concentration of p55 in either supernatant was found in fraction 4.

These results demonstrate that the synthetic Gag expression cassette provides superior production of both p55 protein and VLPs, relative to the native Gag coding sequences.

B. Env Coding Sequences

Human 293 cells were transfected as described in Example 2 with pCMVKm2-based; pCMVlink-based; p-CMVII-based or pESN2-based vectors containing native or synthetic Env expression cassettes. Cells were cultivated for 48 or 60 hours post-transfection. Cell lysates and supernatants were prepared as described (Example 2). Briefly, the cells were washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO)] in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego, CA) were loaded with 20 µl of supernatant or 12.5 µl of cell lysate. A protein molecular weight standard and an HIV SF2 gp120 positive control protein (5 µl, broad size range standard; BioRad Laboratories, Hercules, CA) were also loaded. Electrophoresis was carried out and the proteins were transferred using a BioRad Transfer Chamber (BioRad Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer

(Millipore), where the transfer was performed at 100 volts for 90 minutes. The membranes were then reacted against polyclonal goat anti-gp120SF2/env2-3 anti-sera, followed by incubation with swine anti-goat IgG-peroxidase (POD) (Sigma, St. Louis, MO). Bands indicative of binding were visualized by adding DAB with hydrogen peroxide which deposits a brown precipitate on the membranes.

The results of the immunoblotting analysis showed that cells containing the synthetic Env expression cassette produced the expected Env gp proteins of the predicted molecular weights as determined by mobilities in SDS-polyacrylamide gels at higher per-cell concentrations than cells containing the native expression cassette. The Env proteins were seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants for cells transfected with the synthetic Env expression cassette of the present invention.

20

C. Tat Coding Sequences

Human 293 cells are transfected as described in Example 2 with various vectors containing native or synthetic Tat expression cassettes. Cells are cultivated and isolated proteins analyzed as described above.

Immunoblotting analysis shows that cells containing the synthetic Tat expression cassette produced the expected Tat proteins of the predicted molecular weights as determined by mobilities in SDS-polyacrylamide gels at higher per-cell concentrations than cells containing the native expression cassette.

Example 4Purification of Env polypeptidesA. Purification of Oligomeric gp140

Purification of oligomeric gp140 (o-gp140 US4) was conducted essentially as shown in Figure 60. For the experiments described herein, o-gp140 refers to oligomeric gp140 in either native or modified (e.g., optimized expression sequences, deleted, mutated, truncated, etc.) form. Briefly, concentrated (30-50X) supernatants obtained from CHO cell cultures were loaded onto an anion exchange (DEAE) column which removed DNA and other serum proteins. The eluted material was loaded onto a ceramic hydroxyapatite column (CHAP) which bound serum proteins but not HIV Env proteins. The flow-through from the DEAE and CHAP columns was loaded onto a Protein A column as a precautionary step to remove any remaining serum immunoglobulins. The Env proteins in the flow-through were then captured using the lectin *gluvanthus navalis* (GNA, Vector Labs, Burlingame, CA). GNA has high affinity for mannose rich carbohydrates such as Env. The Env proteins were then eluted with GNA substrate. To remove other highly glycosylated proteins, a cation exchange column (SP) was used to purify gp140/gp120. In a final step, which separates gp120 from o-gp140 , a gel filtration column was used to separate oligomers from monomers. Sizing and chromatography analysis of the final product revealed that this strategy lead to the successful isolation of oligomeric gp140.

30 B. Purification of gp120

Purification of gp120 was conducted essentially as previously described for other Env proteins. Briefly, concentrated supernatants obtained from CHO cell cultures were loaded onto an anion exchange (DEAE) column which

removed DNA and other serum proteins. The eluted material was loaded onto a ceramic hydroxyapatite column (CHAP) which bound serum proteins but not HIV Env proteins. The flow-through from the CHAP column was 5 loaded a cation exchange column (SP) where the flow-through was discarded and the bound fraction eluted with salt. The eluted fraction(s) were loaded onto a Suprose 12/Superdex 200 Tandem column (Pharmacia-Upjohn, Uppsala, Sweden) from which purified gp120 was obtained. Sizing 10 and chromatography analysis of the final product revealed that this strategy successfully purified gp120 proteins.

Example 5

Analysis of Purified Env Polypeptides

15 A. Analysis of o-gp140

It is well documented that HIV Env protein binds to CD4 only in its correct conformation. Accordingly, the ability of o-gp140 US4 polypeptides, produced and purified as described above, to bind CD4 cells was 20 tested. O-gp140 US4 was incubated for 15 minutes with FITC-labeled CD4 at room temperature and loaded onto a Biosil 250 (BioRad) size exclusion column using Waters HPLC. CD4-FITC has the longest retention time (2.67 minutes), followed by CD4-FITC-gp120 (2.167 min). The 25 shortest retention time (1.9 min) was observed for CD4-FITC-o-gp140 US4 indicating that, as expected, o-gp140 US4 binds to CD4 forming a large complex which reduces retention time on the column. Thus, the o-gp140 US4 produced and purified as described above is of the 30 correct size and conformation.

In addition, the US4 o-gp140, purified as described above, was also tested for its ability to bind to a variety of monoclonal antibodies with known epitope specificities for the CD4 binding site, the CD4 inducible

site, the V3 loop and oligomer-specific gp41 epitope. O-gp140 bound strongly to these antibodies, indicating that the purified protein retains its structural integrity.

5 B. Analysis of gp120

As described above, CD4-FITC binds gp120, as demonstrated by the decreased retention time on the HPLC column. Thus, US4 gp120 purified by the above method retains its conformational integrity. In addition, the 10 properties of purified gp120 can be tested by examining its integrity and identity on western blots, as well as, by examining protein concentration, pH, conductivity, endotoxin levels, bioburden and the like. US4 gp120, purified as described above, was also tested for its 15 ability to bind to a variety of monoclonal antibodies with known epitope specificities for the CD4 binding site, the CD4 inducible site, the V3 loop and oligomer-specific gp41 epitope. The pattern of mAb binding to gp120 indicated that the purified protein retained its 20 structural integrity, for example, the purified gp120 did not bind the mAb having the oligomer-specific gp41 epitope (as expected).

Example 6

25 Electron Microscopic Evaluation of VLP Production

The cells for electron microscopy were plated at a density of 50-70% confluence, one day before transfection. The cells were transfected with 10 µg of DNA using transfection reagent LT1 (Panvera) and 30 incubated for 5 hours in serum-reduced medium (see Example 2). The medium was then replaced with normal medium (see Example 2) and the cells were incubated for 14 hours (COS-7) or 40 hours (CHO). After incubation the cells were washed twice with PBS and fixed with 2%

glutaraldehyde. Electron microscopy was performed by Prof. T.S. Benedict Yen, Veterans Affairs, Medical Center, San Francisco, CA).

Electron microscopy was carried out using a 5 transmission electron microscope (Zeiss 10c). The cells were pre-stained with osmium and stained with uranium acetate and lead citrate. The magnification was 100,000X.

Figures 3A and 3B show micrographs of CHO cells 10 transfected with pCMVKM2 carrying the synthetic Gag expression cassette (SEQ ID NO:5) or carrying the Gag-prot expression cassette (SEQ ID NO:79). In the figure, free and budding immature virus-like-particles (VLP) of the expected size (100 nm) are seen for the Gag 15 expression cassette (Figure 3A) and both immature and mature VLPs are seen for the Gag-prot expression cassette (Figure 3B). COS-7 cells transfected with the same vector have the same expression pattern. VLP can also be found intracellularly in CHO and COS-7 cells.

Native and synthetic Gag expression cassettes were 20 compared for their associated levels of VLP production when used to transfect human 293 cells. The comparison was performed by density gradient ultracentrifugation of cell supernatants and Western-blot analysis of the 25 gradient fractions. There was a clear improvement in production of VLPs when using the synthetic Gag construct.

Example 7

30 Expression of Virus-like Particles in the Baculovirus System

A. Expression of Native HIV p55 Gag

To construct the native HIV p55 Gag baculovirus shuttle vector, the prototype SF2 HIV p55 plasmid, pTM1-

Gag (Selby M.J., et al., *J Virol.* 71(10):7827-7831, 1997), was digested with restriction endonucleases NcoI and BamHI to extract a 1.5 Kb fragment that was subsequently subcloned into pAcC4 (*Bio/Technology* 6:47-55, 1988), a derivative of pAc436. Generation of the recombinant baculovirus was achieved by co-transfected 5 2 µg of the HIV p55 Gag pAcC4 shuttle vector with 0.5 µg of linearized, *Autographa californica* baculovirus (AcNPV) wild-type viral DNA into *Spodoptera frugiperda* (Sf9) 10 cells (Kitts, P.A., Ayres M.D., and Possee R.D., *Nucleic Acids Res.* 18:5667-5672, 1990). The isolation of recombinant virus expressing HIV p55 Gag was performed according to standard techniques (O'Reilly, D.R., L.K. 15 Miller, and V. A. Luckow, *Baculovirus Expression Vector: A Laboratory Manual*, W.H. Freeman and Company, New York, 1992).

Expression of the HIV p55 Gag was achieved using a 500 ml suspension culture of Sf9 cells grown in serum-free medium (Miaorella, B., D. Inlow, A. Shauger, and D. 20 Harano, *Bio/Technology* 6:1506-1510, 1988) that had been infected with the HIV p55 Gag recombinant baculovirus at a multiplicity of infection (MOI) of 10. Forty-eight hours post-infection, the supernatant was separated by centrifugation and filtered through a 0.2 µm filter. 25 Aliquots of the supernatant were then transferred to Polyclear™ (Beckman Instruments, Palo Alto, CA) ultracentrifuge tubes, underlaid with 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 24,000 rpm using a Beckman SW28 rotor.

The resulting pellet was suspended in Tris buffer (20 mM Tris HCl, pH 7.5, 250 mM NaCl, and 2.5 mM ethylenediaminetetraacetic acid [EDTA]), layered onto a 20-60% (wt/wt) sucrose gradient, and subjected to 2 hours 30 centrifugation at 40,000 rpm using a Beckman SW41ti

rotor. The gradient was then fractionated starting at the top (20% sucrose) of the gradient into approximately twelve 0.75 ml aliquots. A sample of each fraction was electrophoresed on 8-16% SDS polyacrylamide gels and the resulting bands were visualized after commassie staining (Figure 4). Additional aliquots were subjected to refractive index analysis.

The results shown in Figure 4 indicated that the p55 Gag virus-like particles banded at a sucrose density of range of 1.15 - 1.19 g/ml with the peak at approximately 1.17 g/ml. The peak fractions were pooled and concentrated by a second 20% sucrose pelleting. The resulting pellet was suspended in 1 ml of Tris buffer (described above). The total protein yield as estimated by Bicinchrominic Acid (BCA) (Pierce Chemical, Rockford, IL) was 1.6 mg.

B. Expression of Synthetic HIV p55 Gag

A baculovirus shuttle vector containing the synthetic p55 Gag sequence was constructed as follows. The synthetic HIV p55 expression cassette (Example 1) was digested with restriction enzyme *Sall*I followed by incubation with T4-DNA polymerase. The resulting fragment was isolated (PCR Clean-Up™, Promega, Madison, WI) and then digested with *BamHI* endonuclease. The shuttle vector pAcC13 (Munemitsu S., et al., Mol Cell Biol. 10(11):5977-5982, 1990) was linearized by digestion with *EcoI*, followed by incubation with T4-DNA polymerase, and then isolated (PCR Clean-Up™). The linearized vector was digested with *BamHI*, treated with alkaline phosphatase, and isolated by size fragmentation in an agarose gel. The isolated 1.5 kb fragment was ligated with the prepared pAcC13 vector. The resulting clone was designated pAcC13-Modif.p55Gag.

The expression conditions for the synthetic HIV p55 VLPs differed from those of the native p55 Gag as follows: a culture volume of 1 liter used instead of 500 ml; *Trichoplusia ni* (Tn5) (Wickham, T.J., and Nermrow, 5 G.R., *BioTechnology Progress*, 9:25-30, 1993) insect cells were used instead of Sf9 insect cells; and, an MOI of 3 was instead of an MOI of 10. Experiments performed in support of the present invention showed that there was no appreciable difference in expression level between the 10 Sf9 and Tn5 insect cells with the native p55 clone. In terms of MOI, experience with the native p55 clone suggested that an MOI of 10 resulted in higher expression (approximately 2-fold) of VLPs than a lower MOI.

The sucrose pelleting and banding methods used for 15 the synthetic p55 VLPs were similar to those employed for the native p55 VLPs (described above), with the following exceptions: pelleted VLPs were suspended in 4 ml of phosphate buffered saline (PBS) instead of 1.0 ml of the Tris buffer; and four, 20-60% sucrose gradients were used 20 instead of a single gradient. Also, due to the high concentration of banded VLPs, further concentration by pelleting was not required. The peak fractions from all 4 gradients were simply dialyzed against PBS. The approximate density of the banded VLPs ranged from 1.23- 25 1.28 g/ml. A total protein yield as estimated by BCA was 46 mg. Results from the sucrose gradient banding of the synthetic p55 are shown in Figure 5.

A comparison of the total amount of purified HIV p55 Gag from several preparations obtained from the two 30 baculovirus expression cassettes has been summarized in Figure 6. The average yield from the native p55 was 3.16 mg/liter of culture (n=5, standard deviation (sd) ±1.07, range = 1.8-4.8 mg/L) whereas the average yield from the

synthetic p55 was more than ten-fold higher at 44.5 mg/liter of culture (n=2, sd= \pm 6.4).

In addition to a higher total protein yield, the final product from the synthetic p55-expressed Gag 5 consistently contained lower amounts of contaminating baculovirus proteins than the final product from the native p55-expressed Gag. This difference can be seen in the two commassie-stained gels Figures 4 and 5.

10 C. Expression of Native and Synthetic Gag-Core

Expression of the HIV p55 Gag/HCV Core 173 (SEQ ID NO:8) was achieved using a 2.5 liter suspension culture of Sf9 cells grown in serum-free medium (Miaorella, B., D. Inlow, A. Shauger, and D. Harano. 1988 Bio/Technology 6:1506-1510). The cells were infected with an HIV p55 Gag/HCV Core 173 recombinant baculovirus. Forty-eight hours post-infection, the supernatant was separated from the cells by centrifugation and filtered through a 0.2 μ m filter. Aliquots of the supernatant were then 15 transferred to a Polyclear™ (Beckman Instruments, Palo Alto, CA) ultracentrifuge tubes containing 30% (wt/wt) sucrose, and subjected to 2 hours of centrifugation at 24,000 rpm in a Beckman SW28 rotor and ultracentrifuge.

The resulting pellet was suspended in Tris buffer 25 (50 mM Tris-HCl, pH 7.5, 500 mM NaCl) and layered onto a 30-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor and ultracentrifuge. The gradient was then fractionated starting at the top (30%) of the gradient into 30 approximately 11 x 1.0 ml aliquots. A sample of each fraction was electrophoresed on 8-16% SDS polyacrylamide gels and the resulting bands were visualized after commassie staining.

A subset of aliquots were also subjected to Western blot analysis using monoclonal antibody 76C.5EG (Steimer, K.S., et al., *Virology* 150:283-290, 1986) which is specific for HIV p24 (a subunit of HIV p55). The peak 5 fractions from the sucrose gradient were pooled and concentrated by a second 20% sucrose pelleting. The resulting pellet was suspended in 1 ml of buffer Tris buffer and the total protein yield as estimated by BCA (Pierce Chemical, Rockford, IL) was ~ 1.0 mg.

10 The results from the SDS PAGE are shown in Figure 8 and the anti- p24 Western blot results are shown in Figure 9. Taken together, these results indicate that the HIV p55 Gag/HCV Core 173 chimeric VLPs banded at a sucrose density similar to that of the HIV p55 Gag VLPs 15 and the visible protein band that migrated at a molecular weight of ~ 72,000 kd was reactive with the HIV p24-specific monoclonal antibody. An additional immunoreactive band at approximately 55,000 kd also appeared to be reactive with the anti-p24 antibody and 20 may be a degradation product.

Although aliquots from the above preparation were not tested for reactivity with an HCV Core-specific antibody (an anti-CD22 rabbit serum), results from a similar preparation are shown in Figure 10 and indicate 25 that the main HCV Core-specific reactivity migrates at an approximate molecular weight of 72,000 kd which is in accordance with the predicted molecular weight of the chimeric protein.

The expression conditions for the synthetic HIV p55 30 Gag/HCV Core 173 (SEQ ID NO:8) VLPs differed from those of the native p55 Gag and are as follows: a culture volume of 1 liter used instead of 2.5 liters, *Trichoplusia ni* (Tn5) (Wickham, T.J., and Nemerow, G.R. 1993 *BioTechnology Progress*, 9:25-30) insect cells were

used instead of Sf9 insect cells and an MOI of 3 was instead of an MOI of 10. The sucrose pelleting and banding methods used for the synthetic HIV p55 Gag/HCV Core 173 VLPs were similar to those employed for the native HIV p55 Gag/HCV Core 173 VLPs. However, differences included: pelleted VLPs were suspended in 1 ml of phosphate buffered saline (PBS) instead of 1.0 ml of the Tris buffer, and a single 20-60% sucrose gradients was used. A comparison of the total amount of purified HIV p55 Gag/HCV Core 173 from multiple preparations obtained from the two baculovirus expression cassettes showed that there was an increase in expression using the synthetic HIV p55 Gag/HCV Core 173 cassette.

15 D. Alternative method for the enrichment of HIV p55 Gag VLPs

In addition to purification from the media, p55 (Gag protein) expressed in baculovirus (e.g., using a synthetic expression cassette of the present invention) can also be purified as virus-like particles from the infected insect cells. For example, forty-eight hours post infection, the media and cell pellet are separated by centrifugation and the cell pellet is stored at -70°C until future use. At the time of processing, the cell pellet is suspended in 5 volumes of hypotonic lysis buffer (20 mM Tris-HCl, pH 8.2, 1 mM EGTA; 1 mM MgCl₂, and Complete Protease Inhibitor® (Boehringer Mannheim Corp., Indianapolis, IN]). If needed, the cells are then dounced 8-10 times to complete cell lysis.

The lysate is then centrifuged at approximately 1000-1500 x g for 20 minutes. The supernatant is

decanted into UltraClear™ tubes, underlaid with 20% sucrose (w/w) and centrifuged at 24,000 rpm in SW28 buckets for 2 hours. The resulting pellet is suspended in Tris buffer (20 mM 5 Tris HCl, pH 7.5, 250 mM NaCl, and 2.5 mM ethylene-diamine-tetraacetic acid (EDTA) with 0.1% IGEPAL detergent (Sigma Chemical, St. Louis, MO) and 250 units/ml of benzonase (American International Chemical, Inc., Natick, MA) and 10 incubated at 4°C for at least 30 minutes. The suspension is subsequently layered onto a 20-60% sucrose gradient and spun at 40,000 rpm using an SW41ti rotor for 20-24 hours.

After ultracentrifugation, the sucrose gradient is fractionated and aliquots run on SDS PAGE to identify peak fractions. The peak fractions are dialyzed against PBS and measured for protein content. Negatively stained electron micrographs typically show non-enveloped VLPs 20 somewhat smaller in diameter (80-120 nm) than the budded VLPs. HIV Gag VLPs prepared in this manner are also capable of generating Gag-specific CTL responses in mice.

Example 8

25 In Vivo Immunogenicity of Synthetic Gag Expression

Cassettes

A. Immunization

To evaluate the possibly improved immunogenicity of the synthetic Gag expression cassettes, a mouse study was 30 performed. The plasmid DNA, pCMVKM2 carrying the synthetic Gag expression cassette, was diluted to the following final concentrations in a total injection volume of 100 µl: 20 µg, 2 µg, 0.2 µg, and 0.02 µg. To

overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample was brought up to 20 µg using the vector (pCMVKM2) alone.

As a control, plasmid DNA of the native Gag expression cassette was handled in the same manner. Twelve groups of four Balb/c mice (Charles River, Boston, MA) were intramuscularly immunized (50 µl per leg, intramuscular injection into the tibialis anterior) according to the schedule in Table 7.

10

Table 7

Group	Gag Expression Cassette	Concentration of Gag plasmid DNA (µg)	Immunized at time (weeks):
1	Synthetic	20	0 ¹ , 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	20	0
6	Synthetic	2	0
7	Synthetic	0.2	0
8	Synthetic	0.02	0
9	Native	20	0
10	Native	2	0
11	Native	0.2	0
12	Native	0.02	0

1 = initial immunization at "week 0"

25 Groups 1-4 were bled at week 0 (before immunization), week 4, week 6, week 8, and week 12. Groups 5-12 were bled at week 0 (before immunization) and at week 4.

B. Humoral Immune Response

The humoral immune response was checked with an anti-HIV Gag antibody ELISAs (enzyme-linked immunosorbent assays) of the mice sera 0 and 4 weeks post immunization (groups 5-12) and, in addition, 6 and 8 weeks post immunization, respectively, 2 and 4 weeks post second immunization (groups 1-4).

The antibody titers of the sera were determined by anti-Gag antibody ELISA. Briefly, sera from immunized mice were screened for antibodies directed against the HIV p55 Gag protein. ELISA microtiter plates were coated with 0.2 µg of HIV-1_{SP2} p24-Gag protein per well overnight and washed four times; subsequently, blocking was done with PBS-0.2% Tween (Sigma) for 2 hours. After removal of the blocking solution, 100 µl of diluted mouse serum was added. Sera were tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates were washed four times and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates were washed and 100 µl of 3', 5', 5'-tetramethyl benzidine (TMB; Pierce) was added per well. The optical density of each well was measured after 15 minutes. The titers reported are the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.). The ELISA results are presented in Table 8.

Table 8

Group	Inoculum (μ g)	Expression cassette	Sera - Week 4 ³	Sera - Week 6	Sera - Week 8
1	20	S ¹ - gag	98	455	551
2	2	S - gag	59	1408	227
3	0.	S - gag	29	186	61
4	0.02	S - gag	< 20	< 20	< 20
5	20	S - gag	67	n.a. ⁴	n.a.
6	2	S - gag	63	n.a.	n.a.
7	0.	S - gag	57	n.a.	n.a.
8	0.02	S - gag	< 20	n.a.	n.a.
9	20	N ² - gag	43	n.a.	n.a.
10	2	N - gag	< 20	n.a.	n.a.
11	0.	N - gag	< 20	n.a.	n.a.
15	0.02	N - gag	< 20	n.a.	n.a.

1 = synthetic gag expression cassette (SEQ ID NO: 4)

2 = native gag expression cassette (SEQ ID NO: 1)

3 = geometric mean antibody titer

4 = not applicable

20

The results of the mouse immunizations with plasmid-DNAs show that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression cassettes. Also, the second boost immunization induced a secondary immune response after two weeks (groups 1-3).

25
C. Cellular Immune Response
The frequency of specific cytotoxic T-lymphocytes (CTL) was evaluated by a standard chromium release assay of peptide pulsed Balb/c mouse CD4 cells. Gag expressing vaccinia virus infected CD-8 cells were used as a positive control (vvGag). Briefly, spleen cells (Effector cells, E) were obtained from the BALB/c mice immunized as described above (Table 8) were cultured, restimulated, and assayed for CTL activity against Gag

peptide-pulsed target cells as described (Doe, B., and Walker, C.M., AIDS 10(7):793-794, 1996). The HIV-1_{SP2} Gag peptide used was p7g SEQ ID NO:10. Cytotoxic activity was measured in a standard ⁵¹Cr release assay. Target (T) 5 cells were cultured with effector (E) cells at various E:T ratios for 4 hours and the average cpm from duplicate wells was used to calculate percent specific ⁵¹Cr release. The results are presented in Table 9.

Cytotoxic T-cell (CTL) activity was measured in 10 splenocytes recovered from the mice immunized with HIV Gag DNA (compare Effector column, Table 9, to immunization schedule; Table 8). Effector cells from the Gag DNA-immunized animals exhibited specific lysis of Gag p7g peptide-pulsed SV-BALB (MHC matched) targets cells 15 indicative of a CTL response. Target cells that were peptide-pulsed and derived from an MHC-unmatched mouse strain (MC57) were not lysed (Table 9; MC/p7g).

Table 9

Table 9. Cytotoxic T-lymphocyte (CTL) responses in mice immunized with HIV-1 gag DNA				
Immunization	E:T	Percent specific lysis of target cells*		
		SVBALB none	SVBALB p7g	RMA p7g
20 µg DNA gagmod	100:1	2	49	<1
	30:1	3	30	<1
	10:1	<1	14	<1
2 µg DNA gagmod	100:1	2	37	<1
	30:1	2	21	<1
	10:1	<1	13	<1
0.2 µg DNA gagmod	100:1	2	32	<1
	30:1	3	25	<1
	10:1	1	14	<1
0.02 µg DNA gagmod	100:1	1	17	<1
	30:1	1	16	<1
	10:1	1	8	<1
20 µg DNA gag native	100:1	2	49	<1
	30:1	2	24	<1
	10:1	1	12	<1
2 µg DNA gag native	100:1	<1	18	<1
	30:1	1	14	<1
	10:1	1	7	<1
0.2 µg DNA gag native	100:1	3	30	<1
	30:1	3	17	<1
	10:1	2	7	<1
0.02 µg DNA gag native	100:1	4	2	<1
	30:1	1	2	<1
	10:1	1	2	<1

*representative results of two animals per DNA-dose;

positive CTL responses are indicated by boxed data

The results of the CTL assays show increased potency of synthetic Gag expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

Example 9In vivo Immunization with Env polypeptidesA. Immunogenicity Study of US4 o-gp140 in Ras-3c Adjuvant System

5 Studies have been conducted using rabbits immunized with US4 o-gp140 purified as described above. Studies are also underway in animals to determine immunogenicity of US4 gp120, SF162 o-gp140 and SF162 gp120.

10 Two rabbits (#1 and #2) were immunized intramuscularly at 0, 4, 12 and 24 weeks with 50 µg of US4 o-gp140 in the Ribi™ adjuvant system (RAS-3c), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphoryl lipid A (MPL, Ribi Immunochem, Hamilton, MT).
15 In each experiment described herein, o-gp140 can be native, mutated and/or modified. Antibody responses directed against the US4 o-gp140 protein were measured by ELISA. Results are shown in Table 10.

Table 10

Rabbit/sample	Approximate o-gp140 ELISA titer
pre-immunization	0
#1: post1 (0 week immuniz)	400
#1: post2 (4 week immuniz)	15,000
#1: post3 (12 week immuniz)	50,000
#1: post4 (24 week immuiz)	100,000
#2: post1 (0 week immuniz)	600
#2: post2 (4 week immuniz)	12,000
#2: post3 (12 week immuniz)	25,000
#2: post4 (24 week immuiz)	55,000

The avidities of antibodies directed against the US4 o-gp140 protein were measured in a similar ELISA format employing successive washes with increasing concentrations of ammonium isothiocyanate. Results are shown in Table 11.

Table 11

Time of sample	Approx. Antibody avidity (NH ₄ HCN Conc. in M)
pre-immunization	0.02
post1 (0 week immuniz)	1.8
post2 (4 week immuniz)	3.5
post3 (12 week immuniz)	5.5
post4 (24 week immuniz)	5.1

These results show that US4 o-gp140 is highly immunogenic and able to induce substantial antibody responses after only one or two immunizations.

5 B. Immunogenicity of US4 o-gp140 in MF59-based Adjuvants

Groups of 4 rabbits were immunized intramuscularly at 0, 4, 12 and 24 weeks with various doses of US4 o-gp140 protein in three different MF59-based adjuvants (MF59 is described in International Publication No. WO 90/14837 and typically contains 5% Squalene, 0.5% Tween 80, and 0.5% Span 85). Antibody titers were measured post-third by ELISA using SF2 gp120 to coat the plates. QHC is a quill-based adjuvant (Iscotek, Uppsala, Sweden). Results are shown in Table 12.

15

Table 12

Antigen dose (μ g)	Adjuvant	Anti-gp120 _{SF2} Ab GMT*
12.5	MF59	7231
25	MF59	8896
50	MF59	12822
12.5	MF59/MPL	24146
25	MF59/MPL	27199
50	MF59/MPL	23059
50	MF59/MPL/QHC	31759

*GMT = geometric mean titer

Thus, adjuvanted o-gp140 generated antigen-specific antibodies. Further, the antibodies were shown to increased in avidity over time.

30

C. Neutralizing Antibodies

Neutralizing antibodies post-third immunization were measured against HIV-1 SF2 in a T-cell line adapted virus

(TCLA) assay and against PBMC-grown HIV-1 variants SF2, SF162 and 119 using the CCR5+ CEMx174 LTR-GFP reporter cell line, 5.25 (provided by N. Landau, Salk Institute, San Diego, CA) as target cells. Results are shown in Table 13.

5

Table 13

Neutralizing antibody responses in rabbits immunized
with o-gp140.modUS4 protein

Group	Animal	SF2 TCLA*	SF2 PBMC#	SF162 PBMC#	119 PBMC#
Experiment 1					
o-gp140/ Ras-3c 50 mg	217 218	>640 >640	100% 96	49 37	17 29
Experiment 2					
o-gp140/ MF59 50 mg	792 793 794 795	45 50 59 128	71 87 87 92	39 26 13 15	26 4 0 0
o-gp140/ MF59 + MPL 50 mg	804 805 806 807	173 134 N.D.** 441	91 93 95 100	47 28 49 31	18 4 13 15
o-gp140/MF59 + MPL + QHC 50 mg	808 809 810 811	465 496 >640 92	98 100 101 92	46 44 27 24	40 39 4 37

*TCLA neutralizing antibody titers (50% inhibition).

**Not Determined

* % Inhibition at 1:10 dilution of sera with any detectable non-specific inhibition in pre-bleeds subtracted.

35

The above studies in rabbits indicate that the US4 o-gp140 protein is highly immunogenic. When administered with adjuvant, this protein was able to induce substantial antibody responses after only one or two immunizations.

5 Moreover, the adjuvanted o-gp140 protein was able to generate antigen-specific antibodies which increased in avidity after successive immunizations, and substantial neutralizing activity against T-cell line adapted HIV-1. Neutralizing activity was also observed against PBMC-grown

10 primary HIV strains, including the difficult to neutralize CCR5 co-receptor (R5)-utilizing isolates, SF162 and 119.

Example 10

In Vivo Immunoaenicity of Synthetic Env Expression

15

Cassettes

A. General Immunization Methods

To evaluate the immunogenicity of the synthetic Env expression cassettes, studies using guinea pigs, rabbits, mice, rhesus macaques and baboons were performed. The

20 studies were structured as follows: DNA immunization alone (single or multiple); DNA immunization followed by protein immunization (boost); DNA immunization followed by Sindbis particle immunization; immunization by Sindbis particles alone.

25 B. Humoral Immune Response

The humoral immune response was checked in serum specimens from immunized animals with an anti-HIV Env antibody ELISAs (enzyme-linked immunosorbent assays) at various times post-immunization. The antibody titers of

30 the sera were determined by anti-Env antibody ELISA as described above. Briefly, sera from immunized animals were

screened for antibodies directed against the HIV gp120 or gp140 Env protein. Wells of ELISA microtiter plates were coated

overnight with the selected Env protein and washed four times; subsequently, blocking was done with PBS-0.2% Tween (Sigma) for 2 hours. After removal of the blocking solution, 100 µl of diluted mouse serum was added. Sera were tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates were washed four times and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates were washed and 100 µl of 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) was added per well. The optical density of each well was measured after 15 minutes. Titers are typically reported as the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.).

Example 11

DNA-immunization of Baboons Using Synthetic Gag

Expression Cassettes

A. Baboons

Four baboons were immunized 3 times (weeks 0, 4 and 8) bilaterally, intramuscular into the quadriceps using 1mg pCMVKM2.GagMod.SF2 plasmid-DNA (Example 1). The animals were bled two weeks after each immunization and a p24 antibody ELISA was performed with isolated plasma. The ELISA was performed essentially as described in Example 5 except the second antibody-conjugate was an anti-human IgG, g-chain specific, peroxidase conjugate (Sigma Chemical Co., St. Louis, MD 63178) used at a dilution of 1:500. Fifty µg/ml yeast extract was added to the dilutions of plasma

WO 00/39302

samples and antibody conjugate to reduce non-specific background due to

preexisting yeast antibodies in the baboons. The antibody titer results are presented in Table 14.

Table 14

5

10

15

20

25

30

Immunizati on no.	Weeks	Antigen	wpi ^a / Baboon No.	Ab-titer ^b
1	0	gagmod	0 w/219	< 10
		DNA	0 w/220	< 10
			0 w/221	< 10
			0 w/222	< 10
6			2 wp 1st/219	< 10
			2 wp 1st/220	< 10
			2 wp 1st/221	< 10
			2 wp 1st/222	15
14	4	gagmod	2 wp 4th/219	< 10
		DNA	2 wp 4th/220	88
			2 wp 4th/221	< 10
			2 wp 4th/222	56
30	5	gagmod	2 wp 5th/219	< 10
		DNA	2 wp 5th/220	391
			2 wp 5th/221	237
			2 wp 5th/222	222
46	6	gag VLP	2 wp 6th/219	753
		protein	2 wp 6th/219	4330
			2 wp 6th/219	5000
			2 wp 6th/219	2881

^a wpi = weeks post immunization

^b geometric mean antibody titer

35

In Table 14, pre-bleed data are given as Immunization No. 0; data for bleeds taken 2 weeks post-first immunization are given as Immunization No. 1; data for bleeds taken 2 weeks post-second immunization are given as Immunization No. 2; and, data for bleeds taken 2 weeks post-third immunization are given as Immunization No. 3.

Further, lymphoproliferative responses to p24 antigen were also observed in baboons 221 and 222 two weeks post-fourth immunization (at week 14), and enhanced substantially post-boosting with VLP (at week 44 and 76).
5 Such proliferation results are indicative of induction of T-helper cell functions.

B. Rhesus Macaques

The improved potency of the codon-modified gag expression plasmid observed in mouse and baboon studies was confirmed in rhesus macaques. Four of four macaques had detectable Gag-specific CTL after two or three 1 mg doses of modified gag plasmid. In contrast, in a previous study, only one of four macaques given 1 mg doses of plasmid-DNA encoding the wild-type HIV-1_{sf2} Gag showed strong CTL activity that was not apparent until after the seventh immunization. Further evidence of the potency of the modified gag plasmid was the observation that CTL from two of the four rhesus macaques reacted with three nonoverlapping Gag peptide pools, suggesting that as many as three different Gag peptides are recognized and indicating that the CTL response is polyclonal. Additional quantification and specificity studies are in progress to further characterize the T cell responses to Gag in the plasmid-immunized rhesus macaques. DNA immunization of macaques with the modified gag plasmid did not result in significant antibody responses, with only two of four animals seroconverting at low titers. In contrast, in the same study the majority of macaques in groups immunized with p55Gag protein seroconverted and had strong Gag-specific antibody titers. These data suggest that a prime-boost

strategy (DNA-prime and protein-boost) could be very promising for the induction of a strong CTL and antibody response.

In sum, these results demonstrate that the synthetic
5 Gag plasmid DNA is immunogenic in non-human primates.
When similar experiments were carried out using wild-type
Gag plasmid DNA no such induction of anti-p24 antibodies
was observed after four immunizations.

10

Example 12DNA- and Protein Immunizations of Animals Using Env
Expression Cassettes and PolypeptidesA. Guinea Pigs

Groups comprising six guinea pigs each were
15 immunized intramuscularly at 0, 4, and 12 weeks with
plasmid DNAs encoding the gp120.modUS4, gp140.modUS4,
gp140.modUS4.delV1, gp140.modUS4.delV2,
gp140.modUS4.delV1/V2, or gp160.modUS4 coding sequences
of the US4-derived Env. The animals were subsequently
20 boosted at 18 weeks with a single intramuscular dose of
US4 o-gp140.mut.modUS4 protein in MF59 adjuvant. Anti-
gp120 SF2 antibody titers (geometric mean titers) were
measured at two weeks following the third DNA
immunization and at two weeks after the protein boost.
25 Results are shown in Table 15.

Table 15

Group	GMT post-DNA immuniz.	GMT post-protein boost
5	gp120.modUS4	2098
	gp140.modUS4	190
	gp140.modUS4.delV1	341
	gp140.modUS4.delV2	386
	gp140.modUS4.delV1/V2	664
	gp160.modUS4	235

10

These results demonstrate the usefulness of the synthetic constructs to generate immune responses, as well as, the advantage of providing a protein boost to enhance the immune response following DNA immunization.

15

B. Rabbits

Rabbits were immunized intramuscularly and intradermally using a Bioject needless syringe with plasmid DNAs encoding the following synthetic SF162 Env polypeptides: gp120.modsF162, gp120.modsF162.delV2, gp140.modsF162, gp140.modsF162.delV2, gp140.mut.modsF162, gp140.mut.modsF162.delV2, gp160.modsF162, and gp160.modsF162.delV2. Approximately 1 mg of plasmid DNA (pCMVlink) carrying the synthetic Env expression cassette was used to immunize the rabbits. Rabbits were immunized with plasmid DNA at 0, 4, and 12 weeks. At two weeks after the third immunization all of the constructs were shown to have generated significant antibody titers in the test animals. Further, rabbits immunized with constructs containing deletions of the V2 region

generally generated similar antibody titers relative to rabbits immunized with the companion construct still containing the V2 region.

The nucleic acid immunizations are followed by 5 protein boosting with o-gp140.modSF162.delV2 (0.1 mg of purified protein) at 24 weeks after the initial immunization. Results are shown in Table 16.

Table 16

Group	GMT 2wks post-2nd DNA immunization	GMT 2wks post-3rd DNA immunization	GMT 2wks post-protein boost
gp120.modSF162	4573	5899	26033
gp120.modSF162.delV2	3811	3122	29606
gp140.modSF162	1478	710	12882
gp140.modSF162.delV2	1572	819	11067
gp140.mut.modSF162	1417	788	8827
gp140.mut.modSF162.delV2	1378	1207	13301
gp160.modSF162	23	81	7050
gp160.modSF162.delV2	85	459	11568

All constructs are highly immunogenic and generate substantial antigen binding antibody responses after only 2 immunizations in rabbits.

C. Baboons

Groups of four baboons were immunized intramuscularly with 1 mg doses of DNA encoding different forms of synthetic US4 gp140 (see the following table) at 0, 4, 8, 12, 28, and 44 weeks. The animals were also boosted twice with US4 o-gp140 protein (gp140.mut.modUS4) at 44 and 76 weeks using MF59 as adjuvant. Results are shown in Table 17.

Table 17

Animal	Treatment	2 Wks Post 5th DNA immuniza- tion	2 Wks post 6th DNA (plus o- gp140 prot. immuniz.)	2 Wks post 7th DNA (o-gp140 protein only)
5	CY 215	8.3	446	1813
	CY 216	8.3	433	1236
	CY 217	68	1660	2989
	CY 218	101	2556	1610
Geomean:		26.2	951.4	1812.1
10	CY 219	8.3	8.3	421
	CY 220	8.3	8.3	3117
	CY 221	8.3	954	871
	CY 222	8.3	71	916
Geomean:		8.3	46.5	1011.5
15	CY 223	41.4	10497	46432
	CY 224	8.3	979	470
	CY 225	modUS4	2935	3870
	CY 226	47	1209	4009
Geomean:		68.3	2457.4	4289.6
20	CY 227	8.3	56	5001
	CY 228	8.3	806	1170
	CY 229	8.3	48	3402
	CY 230	8.3	38	6520
GMT*:		8.3	95.3	3375.3

*GMT = geometric mean titer

The results in Table 17 demonstrate the usefulness
 25 of the synthetic constructs to generate immune responses
 in primates such as baboons. In addition, all animals

showed evidence of antigen-specific (Env antigen) lymphoproliferative responses.

D. Rhesus Macaques

5 Two rhesus macaques (designated H445 and J408) were immunized with 1 mg of DNA encoding SF162 gp140 with a deleted V2 region (SF162.gp140.delV2) by intramuscular (IM) and intradermal (ID) routes at 0, 4, 8, and 28 weeks. Approximately 100 µg of the protein encoded by
10 the SF162. gp140mut.delV2 construct was also administered in MF59 by IM delivery at 28 weeks.

ELISA titers are shown in Figure 61. Neutralizing antibody activity is shown Tables 18 and 19.
Neutralizing antibody activity was determined against a
15 variety of primary HIV-1 isolates in a primary lymphocyte or "PBMC-based" assay (see the following tables). Further, the phenotypic co-receptor usage for each of the primary isolates is indicated. As can be seen in the tables neutralizing antibodies were detected against
20 every isolate tested, including the HIV-1 primary isolates (i.e., SF128A, 92US660, 92HT593, 92US657, 92US714, 91US056, and 91US054).

Table 18

	Treatment		Bleed 0	Bleed 1	Bleed 2	
	Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
5	EO 456	25µg 120mod DNA	(None)	8.3	45	309
	EO 457			8.3	254	460
	EO 458			8.3	8.3	93
	EO 459			8.3	43	45
	EO 460			8.3	8.3	274
10	EO 461	25µg 120mod DNA	25µg 120mod DNA	8.3	47	1502
	EO 462			8.3	80	5776
	EO 463			8.3	89	3440
	EO 464			8.3	8.3	3347
	EO 465			8.3	69	1127
15	EO 466	50µg 120mod DNA	(None)	8.3	63	102
	EO 467			8.3	112	662
	EO 468			8.3	94	459
	EO 469			8.3	58	48
	EO 470			8.3	95	355
20	EO 471	50µg 120mod DNA	50µg 120mod DNA	8.3	110	9074
	EO 472			8.3	8.3	4897
	EO 473			8.3	49	4089
	EO 474			8.3	59	5280
	EO 475			8.3	8.3	929
25	EO 476	25µg 120mod DNA	Sindbis/Env	8.3		653
	EO 477			8.3	87	22675
	EO 478			8.3	76	3869
	EO 479			8.3		1004
	EO 480			8.3	71	7080

Table 19

	Treatment		Bleed 0	Bleed 1	Bleed 2
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
EO 481			8.3	8.3	8.3
EO 482			8.3	8.3	8.3
EO 483	Sindbis/Env	(None)	8.3	78	103
EO 484			8.3	8.3	32
EO 485			8.3	76	207
EO 486			8.3	8.3	458
EO 487			8.3	8.3	345
EO 488	Sindbis/Env	Sindbis/Env	8.3	8.3	331
EO 489			8.3	103	111
EO 490			8.3	8.3	5636

Lymphoproliferative activity (LPA) was also determined by antigenic stimulation followed by uptake of ³H-thymidine in these animals and is shown in Table 20. Experiment 1 was performed at 14 weeks post third DNA immunization and Experiment 2 was performed at 2 weeks post fourth DNA immunization using DNA and protein. For gp120ThaiE, gp120SF2 and US4 o-gp140, appropriate background values were used to calculate Stimulation Indices (S.I.; Antigenic stimulation CPM/Background CPM).

10

Table 20

S.I.: Calculated as Ag CPM/Background CPM				
Animal/ exp#	gp120Thai E	gp120 SF2	env2-3SF2	o- gp140US4
J408/#1	2	1	1	5
H445/#1	1	1	1	6
J408/#2	1	1	2	3
H445/#2	0	0	3	2

20 As can be seen by the results presented in Table 20 lymphoproliferative responses to o-gp140.US4 antigen were also in all four animals at both experimental time points. Such proliferation results are indicative of induction of T-helper cell functions.

25 The results presented above demonstrate that the synthetic gp140.modSF162.delV2 DNA and protein are immunogenic in non-human primates.

Example 13

In vitro expression of recombinant Sindbis RNA and DNA containing the synthetic Gag or Env expression cassettes

5 A. Synthetic Gag expression cassettes

To evaluate the expression efficiency of the synthetic Gag expression cassette in Alphavirus vectors, the synthetic Gag expression cassette was subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors. Specifically, a cDNA vector construct for in vitro transcription of Sindbis virus RNA vector replicons (pRSIN-luc; Dubensky, et al., *J Virol.* 70:508-519, 1996) was modified to contain a *PmeI* site for plasmid linearization and a polylinker for insertion of heterologous genes. A polylinker was generated using two oligonucleotides that contain the sites *XhoI*, *PmlI*, *Apal*, *NarI*, *XbaI*, and *NotI* (XPANXNF, SEQ ID NO:17, and XPANXNR, SEQ ID NO:18).

The plasmid pRSIN-luc (Dubensky et al., *supra*) was digested with *XhoI* and *NotI* to remove the luciferase gene insert, blunt-ended using Klenow and dNTPs, and purified from an agarose gel using GeneCleanII (Biol01, Vista, CA). The oligonucleotides were annealed to each other and ligated into the plasmid. The resulting construct was digested with *NotI* and *SacI* to remove the minimal Sindbis 3'-end sequence and A_{40} tract, and ligated with an approximately 0.4 kbp fragment from PKSSIN1-BV (WO 97/38087). This 0.4 kbp fragment was obtained by digestion of pKSSIN1-BV with *NotI* and *SacI*, and purification after size fractionation from an agarose gel. The fragment contained the complete Sindbis virus 3'-end, an A_{40} tract and a *PmeI* site for linearization. This new vector construct was designated SINBVE.

The synthetic HIV Gag coding sequence was obtained from the parental plasmid by digestion with EcoRI, blunt-ending with Klenow and dNTPs, purification with GeneCleanII, digestion with *Sal*I, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Gag coding fragment was ligated into the SINBVE vector that had been digested with *Xho*I and *Pml*I. The resulting vector was purified using GeneCleanII and designated SINBVGag. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVGag and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line as described, for example, in U.S. Patent Numbers 5,843,723 and 5,789,245, and then administered *in vivo* as described..

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) was digested with *Sal*I and *Xba*I, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Gag gene was inserted into the pDCMVSIN-beta-gal by digestion of SINBVGag with *Sal*I and *Xho*I, purification using GeneCleanII of the Gag-containing fragment after agarose gel size fractionation, and ligation. The resulting construct was designated pDSIN-Gag, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells were transfected with recombinant Sindbis vector RNA and DNA, respectively. The supernatants and cell lysates were tested with the Coulter p24 capture ELISA (Example 2).

BHK cells were transfected by electroporation with recombinant Sindbis RNA. The expression of p24 (in ng/ml) is presented in Table 21. In the table, SINGag#1 and 2 represent duplicate measurements, and SIN β gal 5 represents a negative control. Supernatants and lysates were collected 24h post transfection.

Table 21

Construct	Supernatant	Lysate
SIN β gal RNA	0	0
SINGag#1 RNA	7 ng	Max (approx. 1 μ g)
SINGag#2 RNA	1 ng	700 ng

293 cells were transfected using LT-1 (Example 2) 15 with recombinant Sindbis DNA. Synthetic pCMVKM2GagMod.SF2 was used as a positive control. Supernatants and lysates were collected 48h post transfection. The expression of p24 (in ng/ml) is presented in Table 22.

20

Table 22

Construct	Supernatant	Lysate
SINGag DNA	3	30
pCMVKM2.GagMod.SF2 DNA	32	42

The results presented in Tables 21 and 22 30 demonstrate that Gag proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic Gag expression cassette (p55Gag.mod).

B. Synthetic Env expression cassettes

To evaluate the expression efficiency of the 35 synthetic Env expression cassette in Alphavirus vectors,

synthetic Env expression cassettes were subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors as described above for Gag.

The synthetic HIV Env coding sequence was obtained from the parental plasmid by digestion with *SalI* and *XbaI*, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Env coding fragment was ligated into the SINBVE vector that had been digested with *XhoI* and *XbaI*. The resulting vector was purified using GeneCleanII and designated SINBVE_n. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVE_n and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line and administered as described above for Gag.

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) was digested with *SalI* and *XbaI*, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Env gene was inserted into the pDCMVSIN-beta-gal by digestion of SINBVE_n with *XbaI* and *XhoI*, purification using GeneCleanII of the Env-containing fragment after agarose gel size fractionation, and ligation. The resulting construct was designated pDSIN-Env, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells were transfected with recombinant Sindbis vector RNA and DNA, respectively. The supernatants and cell lysates were tested by capture ELISA.

BHK cells were transfected by electroporation with recombinant Sindbis RNA. The expression of Env (in ng/ml) is presented in Table 23. In the table, the Sindbis RNA containing synthetic Env expression cassettes 5 are indicated and β gal represents a negative control. Supernatants and lysates were collected 24h post transfection.

Table 23

	Construct	Supernatant (Neat) ng/ml	Lysate (1:10 dilution) ng/ml
10	β gal RNA	0	0
	gp140.modUS4	726	7147
	gp140.modSF162	3529	7772
	gp140.modUS4.delV1/V2	1738	6526
15	gp140.modUS4.delV2	960	3023
	gp140.modSF162.delV2	2772	3359

293 cells were transfected using LT-1 mediated 20 transfection (PanVera) with recombinant Sindbis DNA containing synthetic expression cassettes of the present invention and β gal sequences as a negative control. Supernatants and lysates were collected 48h post 25 transfection. The expression of Env (in ng/ml) is presented in Table 24.

Table 24

Construct	Supernatant (Neat) ng/ml	Lysate (1:10 dilution) ng/ml
βgal	0	0
gp140.modSF162.delV2	1977	801
5 gp140.modSF162	949	746

The results presented in Tables 23 and 24 demonstrated that Env proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic Env expression cassettes of the present invention.

Example 14

15 A. In vivo Immunization with Gag-containing DNA and/or Sindbis particles

CB6F1 mice were immunized intramuscularly at 0 and 4 weeks with plasmid DNA and/or Sindbis vector RNA-containing particles each containing GagMod.SF2 sequences as indicated in Table 25. Animals were challenged with recombinant vaccinia expressing SF2 Gag at 3 weeks post second immunization (at week 7). Spleens were removed from the immunized and challenged animals 5 days later for a standard ⁵¹C release assay for CTL activity. Values shown in Table 25 indicate the results from the spleens of three mice from each group. The boxed values in Table 25 indicate that all groups of mice receiving immunizations with pCMVKm2.GagMod.SF2 DNA and/or SindbisGagMod.SF2 virus particles either alone or in combinations showed antigen-specific CTL activity.

Table 25

Cytotoxic T-lymphocyte (CTL) responses in mice immunized with HIV-1 gagmod DNA and Sindbis gagmod virus particles				
		Percent specific lysis of target cells*		
	E:T	SVBALB none	SVBALB p7g	RMA p7g
5	Immunization			
	pCMVKm2.GagMod.SF2 DNA ^a at 0, 4 wks	100:1 25:1 6:1	5 5 4	20 20 8
	SindbisGagMod.SF2 virus particles ^b at 0, 4 weeks	100:1 25:1 6:1	10 7 5	49 20 12
10	pCMVKm2.GagMod.SF2 DNA at 0 wks SindbisGagMod.SF2 virus particles at 4 wks	100:1 25:1 6:1	9 7 4	58 42 13
	SindbisGagMod.SF2 virus particles at 4 wks	100:1 25:1	5 4	38 18
	pCMVKm2.GagMod.SF2 DNA at 0 wks	6:1	3	13
15				

^a 20 µg^b 10⁷ particles

20 * Challenge with recombinant vaccinia virus expressing HIV-1SF2 Gag at 3 weeks post second immunization (week 7). Spleens taken 5 days later. Ex vivo CTL assay performed by standard ⁵¹Cr release assay. Values seen represent results from 3 pooled mouse spleens per group

25

B. In vivo Immunization with Env-containing DNA and/or Sindbis particles

30 Balb/C mice were immunized intramuscularly at 0 and 4 weeks(as shown in the following table) with plasmid DNA and/or Sindbis-virus RNA-containing particles each containing gp120.modUS4 sequences. Treatment regimes and antibody titers are shown in Table 26. Antibody titers were determined by ELISA using gp120 SF2 protein to coat the plates.

35

Table 26

	Treatment		Bleed 0	Bleed 1 (8 wks)	Bleed 2 (10 wks)	
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd	
5	EO 456	25µg 120mod DNA	(None)	8.3	45	309
	EO 457			8.3	254	460
	EO 458			8.3	8.3	93
	EO 459			8.3	43	45
	EO 460			8.3	8.3	274
10	EO 461	25µg 120mod DNA	25µg 120mod DNA	8.3	47	1502
	EO 462			8.3	80	5776
	EO 463			8.3	89	3440
	EO 464			8.3	8.3	3347
	EO 465			8.3	69	1127
15	EO 466	50µg 120mod DNA	(None)	8.3	63	102
	EO 467			8.3	112	662
	EO 468			8.3	94	459
	EO 469			8.3	58	48
	EO 470			8.3	95	355
20	EO 471	50µg 120mod DNA	50µg 120mod DNA	8.3	110	9074
	EO 472			8.3	8.3	4897
	EO 473			8.3	49	4089
	EO 474			8.3	59	5280
	EO 475			8.3	8.3	929
25	EO 476	25µg 120mod DNA	Sindbis/Env	8.3		653
	EO 477			8.3	87	22675
	EO 478			8.3	76	3869
	EO 479			8.3		1004
	EO 480			8.3	71	7080
30	EO 481	Sindbis/Env	(None)	8.3	8.3	8.3
	EO 482			8.3	8.3	8.3
	EO 483			8.3	78	103
	EO 484			8.3	8.3	32
	EO 485			8.3	76	207
35	EO 486	Sindbis/Env	Sindbis/Env	8.3	8.3	458
	EO 487			8.3	8.3	345
	EO 488			8.3	8.3	331
	EO 489			8.3	103	111
	EO 490			8.3	8.3	5636

40 As can be seen from the data presented above, all of the mice generally demonstrated substantial immunological responses by bleed number 2. For Env, the best results were obtained using either (i) 50 µg of gp120.modUS4 DNA for the first immunization followed by a second

WO 00/39302

immunization using 50 µg of gp120.modUS4 DNA, or (ii) 25 µg of gp120.modUS4 DNA for the first immunization followed by a second immunization using 10⁷ pfus of Sindbis.

5 The results presented above demonstrate that the Env and Gag proteins of the present invention are effective to induce an immune response using Sindbis vector systems which include the synthetic Env (e.g., gp120.modUS4) or Gag expression cassettes.

10

Example 15

Co-Transfection of Env and Gag as Monocistronic and Bicistronic Constructs

DNA constructs encoding (i) wild-type US4 and SF162 Env polypeptides, (ii) synthetic US4 and SF162 Env polypeptides (gp160.modUS4, gp160.modUS4.delV1/V2, gp160.modSF162, and gp120.modSF162.delV2), and (iii) SF2gag polypeptide (i.e., the Gag coding sequences obtained from the SF2 variant or optimized sequences corresponding to the gagSF2 -- gag.modSF2) were prepared. These monocistronic constructs were co-transfected into 293T cells in a transient transfection protocol using the following combinations: gp160.modUS4; gp160.modUS4 and gag.modSF2; gp160.modUS4.delV1/V2; gp160.modUS4.delV1/V2 and gag.modSF2; gp160.modSF162 and gag.modSF2; gp120.modSF162.delV2 and gag.modSF2; and gag.modSF2 alone.

Further several bicistronic constructs were made where the coding sequences for Env and Gag were under the control of a single CMV promoter and, between the two coding sequences, an IRES (internal ribosome entry site (EMCV IRES); Kozak, M., Critical Reviews in Biochemistry and Molecular Biology 27(45):385-402, 1992; Witherell, G.W., et al., Virology 214:660-663, 1995) sequence was

introduced after the Env coding sequence and before the Gag coding sequence. Those constructs were as follows: gp160.modUS4.gag.modSF2, SEQ ID NO:73 (Figure 61); gp160.modUSF162.gag.modSF2, SEQ ID NO:74 (Figure 62); 5 gp160.modUS4.delV1/V2.gag.modSF2, SEQ ID NO:75 (Figure 63); and gp160.modSF162.delV2.gag.modSF2, SEQ ID NO:76 (Figure 64).

Supernatants from cell culture were filtered through 0.45 μ m filters then ultracentrifuged for 2 hours at 10 24,000 rpm (140,000Xg) in an SW28 rotor through a 20% sucrose cushion. The pelleted materials were suspended and layered on a 20-60% sucrose gradient and spun for 2 hours at 40,000 rpm (285,000Xg) in an SW41Ti rotor. Gradients were fractionated into 1.0 ml samples. A total 15 of 9-10 fractions were typically collected from each DNA transfection group.

The fractions were tested for the presence of the Env and Gag proteins (across all fractions). These results demonstrated that the appropriate proteins were 20 expressed in the transfected cells (i.e., if an Env coding sequence was present the corresponding Env protein was detected; if a Gag coding sequence was present the corresponding Gag protein was detected).

Virus like particles (VLPs) were known to be present 25 through a selected range of sucrose densities. Chimeric virus like particles (VLPs) were formed using all the tested combinations of constructs containing both Env and Gag. Significantly more protein was found in the supernatant collected from the cells transfected with 30 "gp160.modUS4.delV1/V2 and gag.modSF2" than in all the other supernatants.

Western blot analysis was also performed on sucrose gradient fractions from each transfection. The results show that bicistronic plasmids gave lower amounts of VLPs

than the amounts obtained using co-transfection with monocistronic plasmids.

In order to verify the production of chimeric VLPs by these cell lines the following electron microscopic analysis was carried out.

293T cells were plated at a density of 60-70% confluence in 100 mm dishes on the day before transfection. The cells were transfected with 10 µg of DNA in transfection reagent LT1 (Panvera Corporation, 545 10 Science Dr., Madison, WI). The cells were incubated overnight in reduced serum medium (opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium was replaced with 10% fetal calf serum, 2% glutamine in IMDM in the morning of the next day and the cells were incubated for 65 hours. 15 Supernatants and lysates were collected for analysis as described above (see Example 2).

The fixed, transfected 293T cells and purified ENV-GAG VLPs were analyzed by electron microscopy. The cells were fixed as follows. Cell monolayers were washed twice 20 with PBS and fixed with 2% glutaraldehyde. For purified VLPs, gradient peak fractions were collected and concentrated by ultracentrifugation (24,000 rpm) for 2 hours. Electron microscopic analysis was performed by Prof. T.S. Benedict Yen (Veterans Affairs, Medical 25 Center, San Francisco, CA).

Electron microscopy was carried out using a transmission electron microscope (Zeiss 10c). The cells were pre-stained with osmium and stained with uranium acetate and lead citrate. Immunostaining was performed 30 to visualize envelope on the VLP. The magnification was 100,000X.

Figures 65A-65F show micrographs of 293T cells transfected with the following constructs: Figure 65A, gag.modSF2; Figure 65B, gp160.modUS4; Figure 65C,

gp160.modUS4.delV1/V2.gag.modsF2 (bicistronic Env and Gag); Figures 65D and 65E, gp160.modUS4.delV1/V2 and gag.modsF2; and Figure 65F, gp120.modsF162.delV2 and gag.modsF2. In the figures, free and budding immature virus-like-particles (VLPs) of the expected size (approximately 100 nm) decorated with the Env protein were seen. In sum, gp160 polypeptides incorporate into Gag VLPs when constructs were co-transfected into cells. The efficiency of incorporation is 2-3 fold higher when constructs encoding V-deleted Env polypeptides from high synthetic expression cassettes are used.

Although preferred embodiments of the subject invention have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

What Is Claimed Is:

1. An expression cassette, comprising
5 a polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide, wherein the polynucleotide sequence encoding said Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20.

10

2. The expression cassette of claim 1, comprising,
a polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide, wherein the polynucleotide sequence encoding said Gag polypeptide 15 comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9.

3. The expression cassette of claim 1, wherein said 20 polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:4.

4. The expression cassette of claim 1, wherein said 25 polynucleotide sequence further includes a polynucleotide sequence encoding an HIV protease polypeptide.

5. The expression cassette of claim 4, wherein the nucleotide sequence encoding said polypeptide comprises a 30 sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:78, and SEQ ID NO:79.

6. The expression cassette of claim 1, wherein said

polynucleotide sequence further includes a polynucleotide sequence encoding an HIV reverse transcriptase polypeptide.

5 7. The expression cassette of claim 6, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, and SEQ 10 ID NO:84.

15 8. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV tat polypeptide.

20 9. The expression cassette of claim 8, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88 and SEQ ID NO:89.

25 10. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV polymerase polypeptide, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6.

30 11. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV polymerase polypeptide, wherein (i) the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90%

sequence identity to the sequence presented as SEQ ID NO:4, and (ii) wherein the sequence is modified by deletions of coding regions corresponding to reverse transcriptase and integrase.

5

12. The expression cassette of claim 11, wherein said polynucleotide sequence preserves T-helper cell and CTL epitopes.

10 13. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HCV core polypeptide, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:7.

15 14. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59).

20 25 15. The expression cassette of claim 14, wherein said Env polypeptide includes sequences flanking a V1 region but has a deletion in the V1 region itself.

30 16. The expression cassette of claim 15, wherein the polynucleotide sequence encoding the polypeptide comprises the sequence presented as SEQ ID NO:65 (Figure 52 gpl60.modUS4.delV1).

17. The expression cassette of claim 14, wherein

said Env polypeptide includes sequences flanking a V2 region but has a deletion in the V2 region itself.

18. The expression cassette of claim 17, wherein
5 the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:60 (Figure 47); and SEQ ID NO:66 (Figure 53).

19. The expression cassette of claim 17, wherein
10 the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:34 (Figure 20); SEQ ID NO:37 (Figure 24); SEQ ID NO:40 (Figure 27); SEQ ID NO:43 (Figure 30); SEQ ID NO:46 (Figure 33); SEQ ID NO:49 (Figure 36); and SEQ ID NO:76 (Figure 64).

20. The expression cassette of claim 14, wherein
said Env polypeptide includes sequences flanking a V1/V2 region but has a deletion in the V1/V2 region itself.

21. The expression cassette of claim 20, wherein
the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:59 (Figure 46); SEQ ID NO:61 (Figure 48); SEQ ID NO:67 (Figure 54); and SEQ ID NO:75 (Figure 63).

22. The expression cassette of claim 20, wherein
the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:35 (Figure 21); SEQ ID NO:38 (Figure 25); SEQ ID NO:41 (Figure 28); SEQ ID NO:44 (Figure 31); SEQ ID NO:47 (Figure 34) and SEQ ID NO:50 (Figure 37).

23. The expression cassette of claim 14, wherein said Env polypeptide has a mutated cleavage site that prevents the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide.

5

24. The expression cassette of claim 23, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:57 (Figure 44); SEQ ID NO:61 (Figure 48); and SEQ ID NO:63 (Figure 50).

10

25. The expression cassette of claim 23, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34).

20

26. The expression cassette of claim 14, wherein said Env polypeptide includes a gp160 Env polypeptide or a polypeptide derived from a gp160 Env polypeptide.

25

27. The expression cassette of claim 26, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:64 (Figure 51); SEQ ID NO:65 (Figure 52); SEQ ID NO:66 (Figure 53); SEQ ID NO:67 (Figure 54); SEQ ID NO:68 (Figure 55); SEQ ID NO:75 (Figure 63); and SEQ ID NO:73 (Figure 61).

30

28. The expression cassette of claim 26, wherein

the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:48 (Figure 35); SEQ ID NO:49 (Figure 36); SEQ ID NO:50 (Figure 37); SEQ ID NO:76 (Figure 64); and SEQ ID NO:74 (Figure 62).

29. The expression cassette of claim 14, wherein said Env polypeptide includes a gp140 Env polypeptide or a polypeptide derived from a gp140 Env polypeptide.

10

30. The expression cassette of claim 29, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:56 (Figure 43); SEQ ID NO:57 (Figure 44); SEQ ID NO:58 (Figure 45); SEQ ID NO:59 (Figure 46); SEQ ID NO:60 (Figure 47); SEQ ID NO:61 (Figure 48); SEQ ID NO:62 (Figure 49); and SEQ ID NO:63 (Figure 50).

31. The expression cassette of claim 29, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:36 (Figure 23); SEQ ID NO:37 (Figure 24); SEQ ID NO:38 (Figure 25); SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34).

32. The expression cassette of claim 14, wherein said Env polypeptide includes a gp120 Env polypeptide or a polypeptide derived from a gp120 Env polypeptide.

33. The expression cassette of claim 32, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:54 (Figure 41); and SEQ ID NO:55 (Figure 42).

5

34. The expression cassette of claim 32, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:33 (Figure 19); SEQ ID NO:34 (Figure 20); and SEQ ID NO:35 (Figure 21).

10

35. The expression cassette of claim 14, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:55 (Figure 42); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); and SEQ ID NO:68 (Figure 55).

15

36. A recombinant expression system for use in a selected host cell, comprising, an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the selected host cell.

20

37. The recombinant expression system of claim 36, wherein said control elements are selected from the group consisting of a transcription promoter, a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

25

38. The recombinant expression system of claim 36, wherein said transcription promoter is selected from the

group consisting of CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

39. A cell comprising an expression cassette of any
5 of claims 1-35, and wherein said polynucleotide sequence
is operably linked to control elements compatible with
expression in the selected cell.

40. The cell of claim 39, wherein the cell is a
10 mammalian cell.

41. The cell of claim 40, wherein the cell is
selected from the group consisting of BHK, VERO, HT1080,
293, RD, COS-7, and CHO cells.

15 42. The cell of claim 41, wherein said cell is a
CHO cell.

43. The cell of claim 39, wherein the cell is an
20 insect cell.

44. The cell of claim 43, wherein the cell is
either *Trichoplusia ni* (Tn5) or Sf9 insect cells.

25 45. The cell of claim 39, wherein the cell is a
bacterial cell.

46. The cell of claim 39, wherein the cell is a
yeast cell.

30 47. The cell of claim 39, wherein the cell is a
plant cell.

48. The cell of claim 39, wherein the cell is an antigen presenting cell.

49. The cell of claim 48, wherein the lymphoid cell
5 is selected from the group consisting of macrophage,
monocytes, dendritic cells, B-cells, T-cells, stem cells,
and progenitor cells thereof.

50. The cell of claim 39, wherein the cell is a
10 primary cell.

51. The cell of claim 39, wherein the cell is an immortalized cell.

15 52. The cell of claim 39, wherein the cell is a tumor-derived cell.

53. A method for producing a polypeptide including HIV Gag polypeptide sequences, said method comprising,
20 incubating the cells of claim 39, under conditions for producing said polypeptide.

54. A method for producing virus-like particles (VLPs), comprising,
25 incubating the cells of claim 39, under conditions for producing said VLPs.

55. A method for producing a composition of virus-like particles (VLPs), comprising,
30 (a) incubating the cells of claim 39, under conditions for producing said VLPs; and
(b) substantially purifying said VLPs to produce a composition of VLPs.

56. A cell line useful for packaging lentivirus vectors, comprising

suitable host cells that have been transfected with an expression vector containing an expression cassette of 5 any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the host cell.

57. The cell line of claim 56, wherein suitable 10 host cells have been transfected with an expression vector containing the expression cassette of any of claims 1-13.

58. The cell line of claim 56, wherein suitable 15 host cells have been transfected with an expression vector containing the expression cassette of claim 1-3.

59. The cell line of claim 56, wherein suitable 20 host cells have been transfected with an expression vector containing the expression cassette of claim 14-35.

60. A gene delivery vector for use in a Mammalian subject, comprising

25 a suitable gene delivery vector for use in said subject, wherein the vector comprises an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the subject.

30 61. A method of DNA immunization of a subject, comprising,

introducing a gene delivery vector of claim 60 into said subject under conditions that are compatible with expression of said expression cassette in said subject.

62. The method of claim 61, wherein said gene delivery vector is a nonviral vector.

5 63. The method of claim 61, wherein said vector is delivered using a particulate carrier.

10 64. The method of claim 63, wherein said vector is coated on a gold or tungsten particle and said coated particle is delivered to said subject using a gene gun.

65. The method of claim 63, wherein said vector is encapsulated in a liposome preparation.

15 66. The method of claim 61, wherein said vector is a viral vector.

67. The method of claim 66, wherein said viral vector is a retroviral vector.

20 68. The method of claim 67, wherein said viral vector is a lentiviral vector.

69. The method of claim 61, wherein said subject is
25 a mammal.

70. The method of claim 69, wherein said mammal is a human.

30 71. A method of generating an immune response in a subject, comprising

transfected cells of said subject a gene delivery vector of claim 60, under conditions that permit the expression of said polynucleotide and production of said

polypeptide, thereby eliciting an immunological response to said polypeptide.

72. The method of claim 71, wherein said vector is
5 a nonviral vector.

73. The method of claim 72, wherein said vector is delivered using a particulate carrier.

10 74. The method of claim 73, wherein said vector is coated on a gold or tungsten particle and said coated particle is delivered to said vertebrate cell using a gene gun.

15 75. The method of claim 73, wherein said vector is encapsulated in a liposome preparation.

76. The method of claim 71, wherein said vector is a viral vector.

20 77. The method of claim 76, wherein said viral vector is a retroviral vector.

78. The method of claim 77, wherein said viral
25 vector is a lentiviral vector.

79. The method of claim 71, wherein said subject is a mammal.

30 80. The method of claim 79, wherein said mammal is a human.

81. The method of claim 71, wherein said transfecting is done *ex vivo* and said transfected cells

are reintroduced into said subject.

82. The method of claim 71, wherein said transfecting is done *in vivo* in said subject.

5

83. The method of claim 71, where said immune response is a humoral immune response.

10 84. The method of claim 71, where said immune response is a cellular immune response.

15 85. A gene delivery vector comprising an alphavirus vector construct, wherein said alphavirus construct comprises an expression cassette according to any one of claims 1 through 35.

86. The gene delivery vector of claim 85, wherein the alphavirus vector construct is a cDNA vector construct.

20

87. The gene delivery vector of claim 85, wherein the alphavirus comprises a recombinant alphavirus particle preparation.

25

88. The gene delivery vector of claim 85, wherein the vector comprises a eukaryotic layered vector initiation system.

30

89. A method of stimulating an immune response in a subject comprising administering the gene delivery vector of any one of claims 85 through 88 in an amount effective to stimulate an immune response in said subject.

90. The method of claim 89, wherein the gene

delivery vector is administered intramuscularly, intramucosally, intranasally, subcutaneously, intradermally, transdermally, intravaginally, intrarectally, orally or intravenously.

1 / 131

orig.gagSF2

ATGGGTGCGAGAGCGTCGGTATTAGCGGGGGAGAATTAGATAATGGGAAAAATTGGTTAAGGCCAGGGGAAAG

Inact. 1
AAAAAAATATAAGTAAAACATATGATGGCAAGCAGGGAGCTAGAACGATTGCAGTCATCCTGGCTGTTAGAA
G G C C G C C

Inact. 2
ACATCAGAAGGCTGCAGACAAATATTGGACAGCTACAGCCATCCCTCAGACAGGATCAGAAACTTAGATCATTAA
G G C C

Inact. 3
TATAATACAGTAGCAACCCCTTATTGTGTACATCAAAGGATAGATGTAAA
G C C G C G

Inact. 4
GAGGAAGAGCAAAACAAAAGTAAGAAAAGGCACACCAAGCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTC
GTCC G C G

AGCCAAAATTACCCCTATAAGTGCAGAACCTACAGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA
TGGGTAAAAGTAGTAGAAGAAAAGGCTTCAGCCCAGAAGTAATACCCATGTTTCAGCATTATCAGAAGGAGCCACC

Inact. 5
CCACAGAGATTAAACACCATGCTAAACACAGTGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACTATCAAT
G CC G G T G C

GAGGAAGCTGCAGAATGGGATAGAGTGCATCCAGTGCATGCAGGGCTATTGCACCAGGCCAAATGAGAGAACCAAGG
GGAAGTGACATAGCAGGAACACTAGTACCCCTCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCCCAGTA

Inact. 6
GGAGAAATCTATAAAAGATGGATAATCTGGGATTAAATAAAATAGTAAGATGTATAGCCCTACCAGCATTCTGGAC
G C G C G G

Inact. 7
ATAAGACAAGGACCAAGGAACCCCTTAGAGATTATGTAGACCGGTTCTATAAAACTCTAACAGG
CAGGATGTAAAAATTGGATGACAGAACCTTGTGGTCCAAATGCAAACCCAGATTGTAAGA
TTGGGACAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGACCCGCCATAAGCAAGAGTT
TTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAACATAATGATGCAGAGAGGCAATTAGGAACCAAGAAAG
ACTGTTAAGTGTTCATTGTGGCAAGAAGGGCACATAGCCAAAATTGCAGGGCCCTAGGAAAAAGGGCTGTTGG
AGATGTGGAAGGGAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTAGGAAGATCTGGCTTCC
TACAAGGGAGGCCAGGGATTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTGGG
GAGGAGAAAACAACCTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACGTATCCTTAACCTCCCTCAGATCACTC
TTGGCAACGACCCCTCGTCACAATAA

FIG. 1

native HIV-1SF2 gag-protease

2 / 131

From here codon optimization + inactivation (GP1) and (GP2)

ATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAAATGGGAAAAAATCGGTTAAGGCCAGGGGAAAG

Inact. 1
A_nAAAATATAAGTAAACATATAAGTATGGGCAAGCAGGGAGCTAGAACGATTGCAGTCATCCTGGCTGTTAGAA
G G C C G C

Inact. 2
ACATCAGAAGGCTGCAGACAATATTGGGACAGCTACAGCCATCCCTCAGACAGGATCAGAAGAACCTTAGATCATTAA
G G C C

Inact. 3
TATAATACTAGCAACCCCTATTGTGTACATCAAAGGATAGATGTAAGAACACCCAAGGAAGCTTAGAGAAAGATA
C C GC C C G

Inact. 4
GAGGAAGAGCAAAACANAAAGTAAGAAAAAGGCACAGCAAGCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTC
GTCC G C S

AGCCAAAATTACCCCTATACTAGTCAGAACCTACAGGGGAAATGGTACATCAGGCCATATCACCTAGAACCTTAAATGCA
TGGTAAAGTAGTAGAAGAAAAGGTTTCAGCCCAGAAGTAATACCCATGTTTCAGCATTATCAGAAGGAGCCACC

Inact. 5
CCACHAGATTAAACACCATGCTAAACACAGSTGGGGGACATCAAGCAGCCATGCAAATGTTAAAGAGACTATCAAT
G CC G G T G

GAGGAAGCTGCAGAATGGGATAGACTGCATCCAGTCATGCAGGGCTATTGCACCAGGCCAAATGAGAGAACCAAGG

GGAAGTGACATAGCAGGAACACTAGTACCCCTCAGGAACAAATAGGATGGATGACAATAATCCACCTATCCCAGTA

Inact. 6
GGAGAAATCTATAAAAGATGGATAATCCCTGGGATTAATAAAATAGTAAGAAATGTATAGCCCTACCAAGCATTCTGGAC
S C G G G C G C G G

ATAAGACAAGGACCAAGGAACCCCTTAGAGATTATGTAGACCGGTTCTATAAAACTCTAAGAGCCAAAGCTTC

CAGGATGTAAAAATTGGATGACAGAACCTTGTGTCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAGCA
C C G G T

Inact. 7
TTGGGACCGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGACCCGCCATAAAGCAAGAGTT
K C C C G

Inact. 8
TTGGGTGAAGCCATGAGCCAAAGTAACAAATCCAGCTAACATAATGATGCAGAGAGGCAATTGGAAACCAAGAAAG
C G G G G C G G C

Inact. 9
ACTGTTAAGTGTTCATTGTGCAAGAACGGGCACATAGCCAAAATTGCAGGGCCCTAGGAAAGGGCTGTTGG
C C G G C C C C G

AGATGTGGAAGGGAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTGGAAAGATCTGGCTTCC

From here no changes to native sequence (GP1) and (GP2)

TACAAGGGAGGCCAGGGATTTCTCAGAGCAGACCAGAGCCAACAGCCCCACCCAGAAGAGAGCTTCAGGTTGG

GAGGAGAAAACAACCTCCCTCAGAACGAGGCCATAGACAAGGAACGTATCCTTAACCTCCCTCAGATCACTC

From here codon optimization + inactivation (GP1)

TTTGGCAACGACCCCTCGTCACATAAGGATGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATA
G C G G C G G

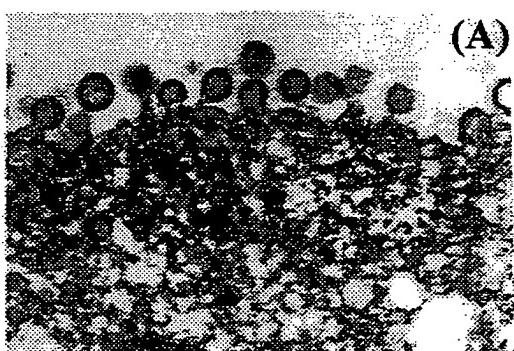
Inact. 12
GACAGTACGATCAGATAACCTGTAGAAATCTGTGGACATAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCA
G G

Inact. 13
ACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGACTTTAAATTCCCCATTAGTCCTATTGAAACTGTACCCAG
C C C C G C C C G

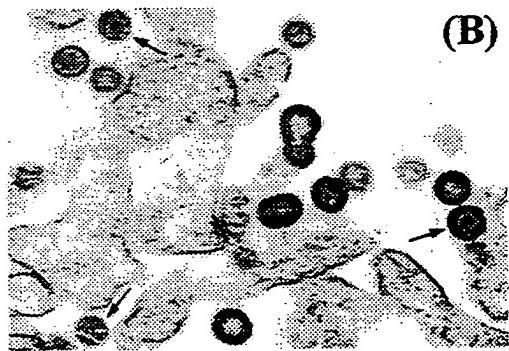
Inact. 14
TAAATTAAGCCAGGAATGGATGGGCCAAAAGTAAAGCAATGCCATTGTAA
G G G G G C C G G

FIG. 2

3 / 131



(A)



(B)

FIG. 3A

FIG. 3B

4 / 131

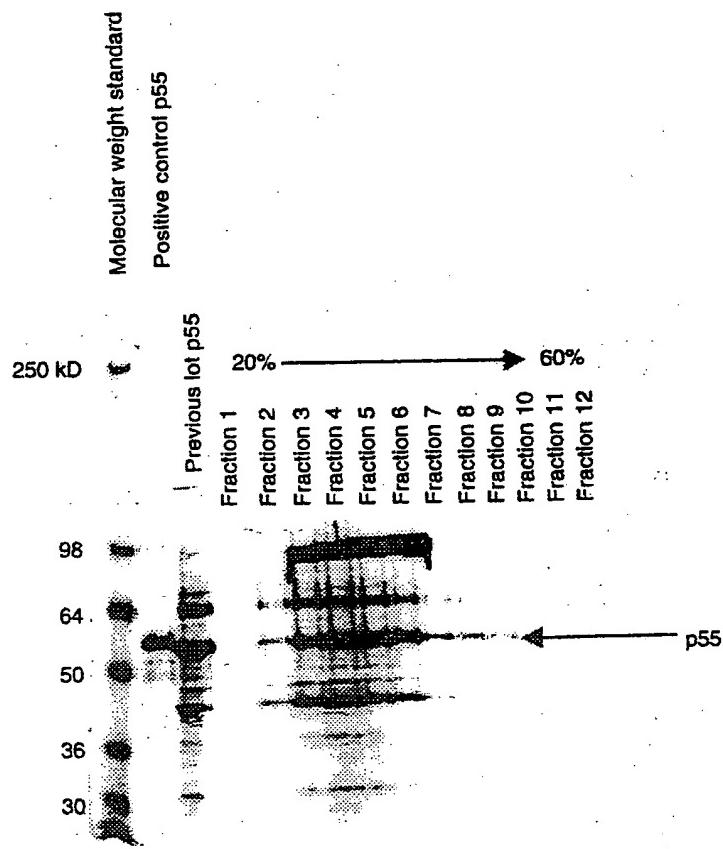


FIG. 4

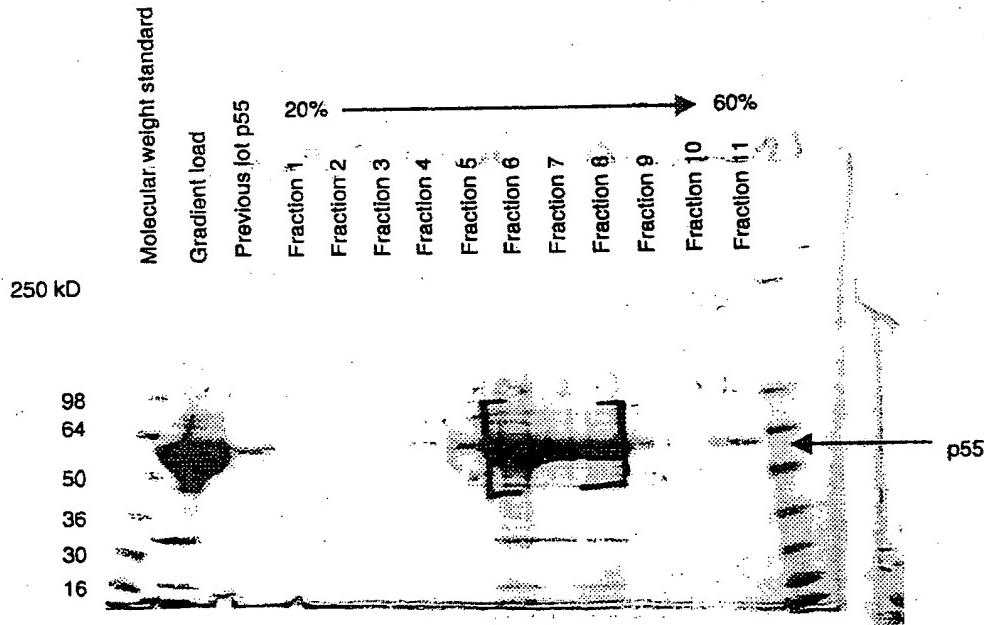


FIG. 5

5 / 131

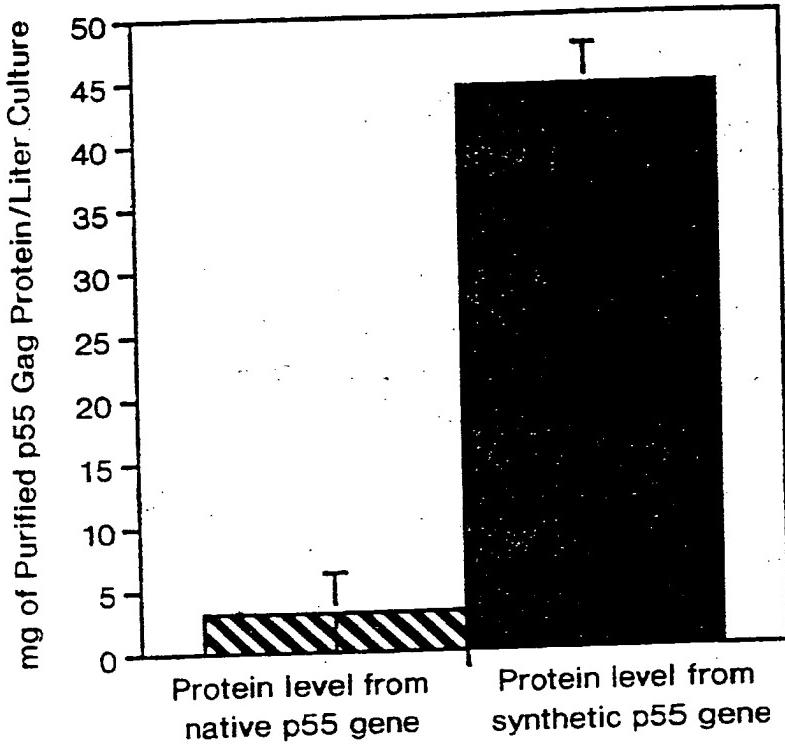


FIG. 6

6 / 131

	50	40	30	20	10	1
GagPol.ModSF	ATGGGGCGCC	GCCCCAGCGT	GCTGAGGCCG	GGCGAGCTGG	ACAACTGGGA	50
GagProt.Mods	ATGGGCGCC	GCCCCAGCGT	GCTGAGGCCG	GGCGAGCTGG	ACAACTGGGA	50
Gag.ModSF2	ATGGGCGCC	GCCCCAGCGT	GCTGAGGCCG	GGCGAGCTGG	ACAACTGGGA	50
GagPol.ModSF	GAAGATCCGC	CTGGGCCCCG	GCGGCAGAAA	GAAGTACAAG	CTGAAAGCACA	100
GagProt.Mods	GAAGATCCGC	CTGGGCCCCG	GCGGCAGAAA	GAAGTACAAG	CTGAAAGCACA	100
Gag.ModSF2	GAAGATCCGC	CTGGGCCCCG	GCGGCAGAAA	GAAGTACAAG	CTGAAAGCACA	100
GagPol.ModSF	TCGTGTGGGC	CAGCCCGGAG	CTGGAGCGCT	TGCGCGTGA	CCCCGGCCTG	150
GagProt.Mods	TCGTGTGGGC	CAGCCCGGAG	CTGGAGCGCT	TGCGCGTGA	CCCCGGCCTG	150
Gag.ModSF2	TCGTGTGGGC	CAGCCCGGAG	CTGGAGCGCT	TGCGCGTGA	CCCCGGCCTG	150
GagPol.ModSF	CTGGAGACCA	GCGAGGGCTG	CCGCCAGATC	CTGGGCCAGC	TGCAAGCCCAG	200
GagProt.Mods	CTGGAGACCA	GCGAGGGCTG	CCGCCAGATC	CTGGGCCAGC	TGCAAGCCCAG	200
Gag.ModSF2	CTGGAGACCA	GCGAGGGCTG	CCGCCAGATC	CTGGGCCAGC	TGCAAGCCCAG	200
GagPol.ModSF	CCTGCAAGACC	GGCAGCGAGG	AGCTGCGAG	CCTGTACAAC	ACCGTGGCCA	250
GagProt.Mods	CCTGCAAGACC	GGCAGCGAGG	AGCTGCGAG	CCTGTACAAC	ACCGTGGCCA	250
Gag.ModSF2	CCTGCAAGACC	GGCAGCGAGG	AGCTGCGAG	CCTGTACAAC	ACCGTGGCCA	250
GagPol.ModSF	CCCTGTACTG	CGTGACCCAG	CGCATCGACG	TCAAGGACAC	CAAGGGGCC	300
GagProt.Mods	CCCTGTACTG	CGTGACCCAG	CGCATCGACG	TCAAGGACAC	CAAGGGGCC	300
Gag.ModSF2	CCCTGTACTG	CGTGACCCAG	CGCATCGACG	TCAAGGACAC	CAAGGGGCC	300
GagPol.ModSF	CTGGGAAAGA	TCGAGGAGGA	GGAGAACAAAG	TCCAAGAAGA	AGGCCAGCA	350
GagProt.Mods	CTGGGAAAGA	TCGAGGAGGA	GGAGAACAAAG	TCCAAGAAGA	AGGCCAGCA	350
Gag.ModSF2	CTGGGAAAGA	TCGAGGAGGA	GGAGAACAAAG	TCCAAGAAGA	AGGCCAGCA	350
GagPol.ModSF	GGCCGCCGCC	GCGCCGGCA	CGGGCAACAG	CAGCCAGGTG	AGCCAGAACT	400
GagProt.Mods	GGCCGCCGCC	GCGCCGGCA	CGGGCAACAG	CAGCCAGGTG	AGCCAGAACT	400
Gag.ModSF2	GGCCGCCGCC	GCGCCGGCA	CGGGCAACAG	CAGCCAGGTG	AGCCAGAACT	400
GagPol.ModSF	ACCCCCATCGT	GCAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	45
GagProt.Mods	ACCCCCATCGT	GCAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	45
Gag.ModSF2	ACCCCCATCGT	GCAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	45

FIG. 7A

7 / 131

GagPol. ModSF	451	CCCCGACCC	TGAACGGCTG	GGTGAAGGTG	GTGAGGAGA	AGGCCTTCAG	500
GagProt. ModS	451	CCCCGACCC	TGAACGGCTG	GGTGAAGGTG	GTGAGGAGA	AGGCCTTCAG	500
Gag. ModSF2	451	CCCCGACCC	TGAACGGCTG	GGTGAAGGTG	GTGAGGAGA	AGGCCTTCAG	500
GagPol. ModSF	501	CCCCGAGGTG	ATCCCCATGT	TCAGGGCCCT	GAGCGAGGGC	GCCACCCCC	550
GagProt. ModS	501	CCCCGAGGTG	ATCCCCATGT	TCAGGGCCCT	GAGCGAGGGC	GCCACCCCC	550
Gag. ModSF2	501	CCCCGAGGTG	ATCCCCATGT	TCAGGGCCCT	GAGCGAGGGC	GCCACCCCC	550
GagPol. ModSF	551	AGGACCTGAA	CACGATGTTG	AACACCGTGG	GGGGCCACCA	GGCGGCCATG	600
GagProt. ModS	551	AGGACCTGAA	CACGATGTTG	AACACCGTGG	GGGGCCACCA	GGCGGCCATG	600
Gag. ModSF2	551	AGGACCTGAA	CACGATGTTG	AACACCGTGG	GGGGCCACCA	GGCGGCCATG	600
GagPol. ModSF	601	CAGATGCTGA	AGAGAACCAT	CAACGAGGAG	GCCGGCCGAGT	GGGACCCGGT	650
GagProt. ModS	601	CAGATGCTGA	AGAGAACCAT	CAACGAGGAG	GCCGGCCGAGT	GGGACCCGGT	650
Gag. ModSF2	601	CAGATGCTGA	AGAGAACCAT	CAACGAGGAG	GCCGGCCGAGT	GGGACCCGGT	650
GagPol. ModSF	651	GCACCCCGTG	CACGGCGGCC	CCATCGCCCC	CGGCCAGATG	CGCGAGCCCC	700
GagProt. ModS	651	GCACCCCGTG	CACGGCGGCC	CCATCGCCCC	CGGCCAGATG	CGCGAGCCCC	700
Gag. ModSF2	651	GCACCCCGTG	CACGGCGGCC	CCATCGCCCC	CGGCCAGATG	CGCGAGCCCC	700
GagPol. ModSF	701	GCGGAGCGA	CATCGCCGGC	ACCACAGCA	CCCTGCAGGA	GCAGATCGGC	750
GagProt. ModS	701	GCGGAGCGA	CATCGCCGGC	ACCACAGCA	CCCTGCAGGA	GCAGATCGGC	750
Gag. ModSF2	701	GCGGAGCGA	CATCGCCGGC	ACCACAGCA	CCCTGCAGGA	GCAGATCGGC	750
GagPol. ModSF	751	TGGATGACCA	ACAACCCCCC	CATCCCCGTG	GGGGAGATCT	ACAAGGGGTG	800
GagProt. ModS	751	TGGATGACCA	ACAACCCCCC	CATCCCCGTG	GGGGAGATCT	ACAAGGGGTG	800
Gag. ModSF2	751	TGGATGACCA	ACAACCCCCC	CATCCCCGTG	GGGGAGATCT	ACAAGGGGTG	800
GagPol. ModSF	801	GATCATCCTG	GGCCTGAACA	AGATCGTGGC	GATGTACAGC	CCCACCAAGCA	850
GagProt. ModS	801	GATCATCCTG	GGCCTGAACA	AGATCGTGGC	GATGTACAGC	CCCACCAAGCA	850
Gag. ModSF2	801	GATCATCCTG	GGCCTGAACA	AGATCGTGGC	GATGTACAGC	CCCACCAAGCA	850
GagPol. ModSF	851	TCCTGGACAT	CGCCAGGGC	CCCAAGGAGC	CCTTCCGGCA	CTACGGGAC	900
GagProt. ModS	851	TCCTGGACAT	CGCCAGGGC	CCCAAGGAGC	CCTTCCGGCA	CTACGGGAC	900
Gag. ModSF2	851	TCCTGGACAT	CGCCAGGGC	CCCAAGGAGC	CCTTCCGGCA	CTACGGGAC	900

FIG. 7B

8 / 131

GagPol . ModSF	910	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCAGGCCAGG	ACGTGAAGAA	950
GagProt . ModS	901	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCAGGCCAGG	ACGTGAAGAA	950
Gag . ModSF2	901	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCAGGCCAGG	ACGTGAAGAA	950
GagPol . ModSF	960	970	980	990	1000		
GagProt . ModS	951	CTGGATGACC	GAGACCCCTGCG	TGGTGCAGAA	CGCCAACCCC	GAATGCAAGA	10000
Gag . ModSF2	951	CTGGATGACC	GAGACCCCTGCG	TGGTGCAGAA	CGCCAACCCC	GAATGCAAGA	10000
GagPol . ModSF	951	CCATCTGAA	GGCTCTCGGC	CCCGGGCCA	CCCTGGAGGA	GGCTGAAGA	10000
GagProt . ModS	1001	CCATCTGAA	GGCTCTCGGC	CCCGGGCCA	CCCTGGAGGA	GGCTGAAGA	10000
Gag . ModSF2	1001	CCATCTGAA	GGCTCTCGGC	CCCGGGCCA	CCCTGGAGGA	GGCTGAAGA	10000
GagPol . ModSF	1010	1020	1030	1040	1050		
GagProt . ModS	1051	GCCTGCCAGG	GCCTGCCAGG	CCCCGGCCAC	AAAGGCCCGG	GATGATGACCC	1050
Gag . ModSF2	1051	GCCTGCCAGG	GCCTGCCAGG	CCCCGGCCAC	AAGGGCCCGG	GATGATGACCC	1050
GagPol . ModSF	1051	GGCGATGAGC	CAGGTGACGA	ACCCGGGCAC	CATCATGATG	CAGCGGGCA	1100
GagProt . ModS	1101	GGCGATGAGC	CAGGTGACGA	ACCCGGGCAC	CATCATGATG	CAGCGGGCA	1100
Gag . ModSF2	1101	GGCGATGAGC	CAGGTGACGA	ACCCGGGCAC	CATCATGATG	CAGCGGGCA	1100
GagPol . ModSF	1110	1120	1130	1140	1150		
GagProt . ModS	1110	1120	1130	1140	1150		
Gag . ModSF2	1110	1120	1130	1140	1150		
GagPol . ModSF	1151	ACTTCGCAA	CAAGCGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
GagProt . ModS	1151	ACTTCGCAA	CAAGCGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
Gag . ModSF2	1151	ACTTCGCAA	CAAGCGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
GagPol . ModSF	1210	1220	1230	1240	1250		
GagProt . ModS	1201	GGCACACCCG	CCAGGAACCTG	CCGGCCCC	CGCAAGAAGG	GCTGCTGGCG	1250
Gag . ModSF2	1201	GGCACACCCG	CCAGGAACCTG	CCGGCCCC	CGCAAGAAGG	GCTGCTGGCG	1250
GagPol . ModSF	1260	1270	1280	1290	1300		
GagProt . ModS	1251	CTGGGCCGC	GAAGGACACC	AAATGAAGA	TGGCACTGAG	AGACAGGGCTA	1300
Gag . ModSF2	1251	CTGGGCCGC	GAAGGACACC	AAATGAAGA	TGGCACTGAG	AGACAGGGCTA	1300
GagPol . ModSF	1301	ATTTTTAGG	GAAGATCTG	CCTTCCTACA	AGGGAAATT	AGGGAAATT	1350
GagProt . ModS	1301	ATTTTTAGG	GAAGATCTG	CCTTCCTACA	AGGGAAATT	AGGGAAATT	1350
Gag . ModSF2	1301	ACTTCCTGG	CAAGATCTG	CCCAAGCTACA	AGGGCCGCC	CGGCAACTTC	1350

FIG. 7C

9 / 131

GagPol. ModSF	1351	CTTCAGAGCA	GACCAGAGCC	AACAGCCCCA	CCAGAAAGAA	GCTTCAGGTT	1400
GagProt. Mods	1351	CTTCAGAGCA	GACCAGAGCC	AACAGCCCCA	CCAGAAAGAA	GCTTCAGGTT	1400
Gag. ModSF2	1351	CTGCAGAGCC	GCCTGGAGCC	CACCGCCCCC	CCCGAGGAGA	GCTTCAGGTT	1400
GagPol. ModSF	1401	TGGGGAGGAG	AAAACAACTC	CCTCTAGAA	GCAGGAGCCG	ATAGACAAGG	1450
GagProt. Mods	1401	TGGGGAGGAG	AAAACAACTC	CCTCTAGAA	GCAGGAGCCG	ATAGACAAGG	1450
Gag. ModSF2	1401	CGGGGAGGAG	AAAAGACCACCC	CCAGGCCAGAA	GCAGGAGCCC	ATCGACAAGG	1450
GagPol. ModSF	1451	AACTGTATCC	TTTAACCTTC	CTCAGATCAC	TCTTTGGCAA	CGACCCCTCG	1500
GagProt. Mods	1451	AACTGTATCC	TTTAACCTTC	CTCAGATCAC	TCTTTGGCAA	CGACCCCTCG	1500
Gag. ModSF2.	1451	AGCTGTACCC	CTTGACCAAGC	CTGCAGGCC	TGTTGGCAA	CGACCCCTCG	1500
GagPol. ModSF	1501	TCACAGTAAG	GATCGGGGGC	CAGCTCAAGG	AGGCCTGCT	CGACACCGGC	1550
GagProt. Mods	1501	TCACAGTAAG	GATCGGGGGC	CAGCTCAAGG	AGGCCTGCT	CGACACCGGC	1550
Gag. ModSF2	1501	AGCCAGTAA.	1550
GagPol. ModSF	1551	GCCGACGACA	CCGTGCTGGA	GGAGATGAAC	CTGCCCCGGCA	AGTGGAGCC	1600
GagProt. Mods	1551	GCCGACGACA	CCGTGCTGGA	GGAGATGAAC	CTGCCCCGGCA	AGTGGAGCC	1600
Gag. ModSF2	1551	1600
GagPol. ModSF	1601	CAAGATGATC	GGCGGGATCG	GGGGCTTCAT	CAAGGTGCGG	CAGTACGACC	1650
GagProt. Mods	1601	CAAGATGATC	GGCGGGATCG	GGGGCTTCAT	CAAGGTGCGG	CAGTACGACC	1650
Gag. ModSF2	1601	1650
GagPol. ModSF	1610	1620	1630	1640	1650	1600	1600
GagProt. Mods	1610	1620	1630	1640	1650	1600	1600
Gag. ModSF2	1610	1600
GagPol. ModSF	1651	AGATCCCCGT	GGAGATCTGC	GGCCACAAAGG	CCATCGGCCAC	CGTGCTGCTG	1700
GagProt. Mods	1651	AGATCCCCGT	GGAGATCTGC	GGCCACAAAGG	CCATCGGCCAC	CGTGCTGCTG	1700
Gag. ModSF2	1651	1700
GagPol. ModSF	1701	GGCCCCACCC	CCGTGAACAT	CATCGGCCGC	AACCTGCTGA	CCCATGTCGG	1750
GagProt. Mods	1701	GGCCCCACCC	CCGTGAACAT	CATCGGCCGC	AACCTGCTGA	CCCATGTCGG	1750
Gag. ModSF2	1701	1750
GagPol. ModSF	1751	CTGCACCCCTG	AACTTCCCCA	TCAGCCCCAT	CGAGACGGTG	CCCGTGAAGC	1800
GagProt. Mods	1751	CTGCACCCCTG	AACTTCCCCA	TCAGCCCCAT	CGAGACGGTG	CCCGTGAAGC	1800
Gag. ModSF2	1751	1800

FIG. 7D

10/131

GagPol.ModSF	1801	TGAAGCCGGG	GATGGACGGC	CCCAAGGTCA	AGCAGTGGCC	CCTGACCGAG	1850
GagProt.ModS	1801	TGAAGCCGGG	GATGGACGGC	CCCAAGGTCA	AGCAGTGGCC	CCTGTAA...	1850
Gag.ModSF2	1801	1860	1870	1880	1890	1900
GagPol.ModSF	1851	GAGAAGATCA	AGGCCCTGGT	GGAGATCTGC	ACCGAGATGG	AGAAGGGGG	1900
GagProt.ModS	1851	1900
Gag.ModSF2	1851	1920	1930	1940
GagPol.ModSF	1901	CAAGATCAGC	AAAGATCGGC	CCGAGAACCC	CTAACACACC	CCCGTGTCG	1950
GagProt.ModS	1901	1950
Gag.ModSF2	1901	1960	1970	1980	1990
GagPol.ModSF	1951	CCATCAAGAA	GAAGGACAGC	ACCAAGTGGC	GCAAGCTGGT	GGACTTCGGC	2000
GagProt.ModS	1951	2000
Gag.ModSF2	1951	2010	2020	2030	2040
GagPol.ModSF	2001	GAGCTAACAA	AGCGCACCCA	GGACTTCTGG	GAGGTGCAGC	TGGGCATCCC	2050
GagProt.ModS	2001	2050
Gag.ModSF2	2001	2060	2070	2080	2090
GagPol.ModSF	2051	CCACCCGGCC	GGCCTGAANGA	AGAAGAAAG	CGTGACCGTG	C'TGGACCGTG	2100
GagProt.ModS	2051	2100
Gag.ModSF2	2051	2110	2120	2130	2140
GagPol.ModSF	2101	GCGGAGCCTA	CTTCAGCGGTG	CCCTGGACA	AGGACTTCGG	CAAGTACACC	2150
GagProt.ModS	2101	2150
Gag.ModSF2	2101	2160	2170	2180	2190
GagPol.ModSF	2151	GCCTTCACCA	TCCCCAGCAT	CAACAACGAG	ACCCCGGCA	TCCGCTACCA	2200
GagProt.ModS	2151	2200
Gag.ModSF2	2151	2210	2220	2230	2240
GagPol.ModSF	2201	GTACAAACGTG	CTGCCCGAGG	GCTGGAAGGG	CAGCCCCGGC	ATCTTCCAGA	2250
GagProt.ModS	2201	2250
Gag.ModSF2	2201	2250

FIG. 7E

11 / 131

GagPol.ModSF	2251	GCAGCATGAC	CAAGATCCTG	GAGCCCTTCC	GCAGCGAGAA	CCCCGACATC	2300
GagProt.ModS	2251	2300
Gag.ModSF2	2251	2300
GagPol.ModSF	2301	GTGATCTACC	AGTACATGGA	CGACCTGTAC	GTGGGCAGGC	ACCTGGAGAT	2350
GagProt.ModS	2301	2350
Gag.ModSF2	2301	2350
GagPol.ModSF	2351	CGGCCAGCAC	CGCACCAAGA	TCGAGGAGCT	GGGCCAGCAC	CTGCTGGGCT	2400
GagProt.ModS	2351	2400
Gag.ModSF2	2351	2400
GagPol.ModSF	2401	GGGGCTTCAC	CACCCCCGAC	AAGAAGCACC	AGAAGGAGCC	CCCCCTTCCTG	2450
GagProt.ModS	2401	2450
Gag.ModSF2	2401	2450
GagPol.ModSF	2451	TGGATGGCT	ACGAGCTGCA	CCCCGACAAG	TGGACCGTGC	AGCCCACAT	2500
GagProt.ModS	2451	2500
Gag.ModSF2	2451	2500
GagPol.ModSF	2501	GCTGCCGAG	AGGACAGCT	GGACCGTGA	CGACATCCAG	AGCTGGTGG	2550
GagProt.ModS	2501	2550
Gag.ModSF2	2501	2550
GagPol.ModSF	2551	GCAAGCTGAA	CTGGGCCAGC	CAGATCTACG	CCGGCATCAA	GGTGANGCAG	2600
GagProt.ModS	2551	2600
Gag.ModSF2	2551	2600
GagPol.ModSF	2601	CTGTGCAAGC	TGCTGGCGGG	CACCAAGGGCC	CTGACCGAGG	TGATCCCCCT	2650
GagProt.ModS	2601	2650
Gag.ModSF2	2601	2650
GagPol.ModSF	2651	GACCGAGGAG	GCCGAGCTGG	AGCTGGCCGA	GAACCGCGAG	ATCCCTGAAGG	2700
GagProt.ModS	2651	2700
Gag.ModSF2	2651	2700

FIG. 7F

12 / 131

GagPol.ModsF	2701	AGCCCCGTGCA	CGAGGTGTAC	TACGACCCA	GCAGGGACCT	GGTGGCCGAG	2750
GagProt.Mods	2701	2750
Gag.ModsF2	2701	2750
GagPol.ModsF	2751	ATCCAGAAC	AGGCCAGGG	CCAGTGAGCC	TACAGATCT	ACCAAGGAGCC	2800
GagProt.Mods	2751	2800
Gag.ModsF2	2751	2800
GagPol.ModsF	2801	CTTCAAAGAAC	CTGAAGACCG	GCAAGTACGC	CCGCATGCGC	GGCGCCACAA	2850
GagProt.Mods	2801	2850
Gag.ModsF2	2801	2850
GagPol.ModsF	2851	CCAACGACGT	GAAGCAGCTG	ACCGAGGCCG	TGCAGAGGT	GAGCACCGAG	2900
GagProt.Mods	2851	2900
Gag.ModsF2	2851	2900
GagPol.ModsF	2901	AGCATGTGA	TCTGGGGCAA	GATCCCCAAG	TTCAGCTGC	CCATCCAGAA	2950
GagProt.Mods	2901	2950
Gag.ModsF2	2901	2950
GagPol.ModsF	2951	GGAGACCTGG	GAGGCCTGGT	GGATGGAGTA	CTGGCAGGGCC	ACCTGGATCC	3000
GagProt.Mods	2951	3000
Gag.ModsF2	2951	3000
GagPol.ModsF	3001	CCGAGTGGGA	GTTCTGTAAAC	ACCCCCCCC	TEGTGAAGCT	GTGGTACCAAG	3050
GagProt.Mods	3001	3050
Gag.ModsF2	3001	3050
GagPol.ModsF	3051	CTGGAGAAGG	AGCCCCATCGT	GGGGCCGGAG	ACCTTCTACG	TGGACGGGCG	3100
GagProt.Mods	3051	3100
Gag.ModsF2	3051	3100
GagPol.ModsF	3101	CGCCAAACCGC	GAGACCAGGC	TGGCAAGGC	GGCTACGTG	ACCGACCGCG	3150
GagProt.Mods	3101	3150
Gag.ModsF2	3101	3150

FIG. 7G

13 / 131

GagPol.	ModSF	3151	GCCGCCAGAA	GGTGGTGAGC	ATCGCCGACA	CCACCAACCA	GAAGACCGAG	3200
GagProt.	ModS	3151	3200
Gag.	ModSF2	3151	3200
GagPol.	ModSF	3201	CTGCAGGCCA	TCCACCTGGC	CCTGCAGGAC	AGGGCCCTGG	AGGTGAACAT	3250
GagProt.	ModS	3201	3250
Gag.	ModSF2	3201	3250
GagPol.	ModSF	3251	CGTGACCGAC	AGCACGTTACG	CCCTGGCAT	CATCCAGGCC	CAGCCCCGACA	3300
GagProt.	ModS	3251	3300
Gag.	ModSF2	3251	3300
GagPol.	ModSF	3301	AGAGGGAGAG	CGAGCTGGTG	AGCCAGATCA	TCGAGCAGCT	GATCAAGAAC	3350
GagProt.	ModS	3301	3350
Gag.	ModSF2	3301	3350
GagPol.	ModSF	3351	GAGAAGGTGT	ACCTGGCCTG	GGTGGCCGCC	CACAAGGGCA	TGGGGCCAA	3400
GagProt.	ModS	3351	3400
Gag.	ModSF2	3351	3400
GagPol.	ModSF	3401	CGAGCAGGTG	GACAAGCTGG	TGAGCCGGGG	CATCCGCAAG	GTGCTGTCC	3450
GagProt.	ModS	3401	3450
Gag.	ModSF2	3401	3450
GagPol.	ModSF	3451	TGAACGGCAT	CGACCAAGGCC	CAGGGAGGC	ACGNGAAGTA	CCACAGGAAAC	3500
GagProt.	ModS	3451	3500
Gag.	ModSF2	3451	3500
GagPol.	ModSF	3501	TGGCGGGCCA	TGGCCAGGCA	CTTCACACTG	CCCCCGTGG	TGGCCAAGGA	3550
GagProt.	ModS	3501	3550
Gag.	ModSF2	3501	3550
GagPol.	ModSF	3551	GATCGTGGCC	AGCTGGCACA	AGTGCAGCT	GAAGGGCGAG	GCCATGGCACG	3600
GagProt.	ModS	3551	3600
Gag.	ModSF2	3551	3600

FIG. 7H

14 / 131

GagPol.ModSF	3601	GCCAGGTGGA	CTGCAGCCCC	GGCATCTGGC	AGCTGGACTG	CACCCACCTG	3650
GagProt.ModS	3601	3650
Gag.ModSF2	3601	3660	3670	3680	3690	3700
Gag.Pol.ModSF	3651	GAGGGCAAGA	TCAATCCTGGT	GGCCGTGCAC	GTGGCCAGCG	GCTACATTCGA	3700
GagProt.ModS	3651	3700
Gag.ModSF2	3651	3710	3720	3730	3740	3750
GagPol.ModSF	3701	GGCCCAGGTG	ATCCCCGCCG	AGACCGGCCA	GGAGAGCCGC	TACTTCTCTGC	3750
GagProt.ModS	3701	3750
Gag.ModSF2	3701	3760	3770	3780	3790	3800
GagPol.ModSF	3751	TGAAGCTGGC	CGGCCGGCTGG	CCCGTGAAAGA	CCATCCACAC	CGACACACGGC	3800
GagProt.ModS	3751	3800
Gag.ModSF2	3751	3810	3820	3830	3840	3850
GagPol.ModSF	3801	AGCAACTTCA	CCAGGACACAC	CGTGAAGGCC	GCCTGCTGGT	GGGGGGCAT	3850
GagProt.ModS	3801	3850
Gag.ModSF2	3801	3860	3870	3880	3890	3900
GagPol.ModSF	3851	CAAGCAGGAG	TTCGGCATCC	CCTACAAACCC	CCAGGCCAG	GGCGTGGTGG	3900
GagProt.ModS	3851	3900
Gag.ModSF2	3851	3910	3920	3930	3940	3950
GagPol.ModSF	3901	AGAGCATGAA	CAACGAGCTG	AAGAAAGATCA	TGGGCCAGGT	GGGCACCCAG	3950
GagProt.ModS	3901	3950
Gag.ModSF2	3901	3960	3970	3980	3990	4000
GagPol.ModSF	3951	GCCGAGGCC	TGAAGACCCG	CGTGCAGATG	GCCGTGTTCGA	TCCACAACTT	4000
GagProt.ModS	3951	4000
Gag.ModSF2	3951	4010	4020	4030	4040	4050
GagPol.ModSF	4001	CAAGCGCAAG	GGGGCATCG	GGGGTACAG	GGCGGGCAAG	CGCAATCTGG	4050
GagProt.ModS	4001	4050
Gag.ModSF2	4001	4050

FIG. 71

15 / 131

GagPol.ModSF	4051	ACATCATCGC	CACCGACATC	CAGACCAGG	AGCTGCAGAA	GCAGATCACCC	4100
GagProt.ModS	4051	4100
Gag.ModSF2	4051	4100
GagPol.ModSF	4101	AAGATCCAGA	ACTTCCGGCT	GTACTACCGC	GACAACAAAGG	ACCCCTGTG	4150
GagProt.ModS	4101	4150
Gag.ModSF2	4101	4150
GagPol.ModSF	4151	GAAGGGCCCC	GCCAAAGCTGC	TGTGAAAGGG	CGAGGGGCC	GTGGTGATCC	4200
GagProt.ModS	4151	4200
Gag.ModSF2	4151	4200
GagPol.ModSF	4201	AGGACAAACAG	CGACATCAG	GTGGTCCCC	GCCGCAAGGC	CAAGATCATC	4250
GagProt.ModS	4201	4250
Gag.ModSF2	4201	4250
GagPol.ModSF	4251	CGCGACTACG	GCAAGCAGAT	GGCCGGCGAC	GAATGCGTGG	CCAGCCGCCA	4300
GagProt.ModS	4251	4300
Gag.ModSF2	4251	4300
GagPol.ModSF	4301	GGACGAGGAC	TAG	4350
GagProt.ModS	4301	4350
Gag.ModSF2	4301	4350

FIG. 7J

16 / 131

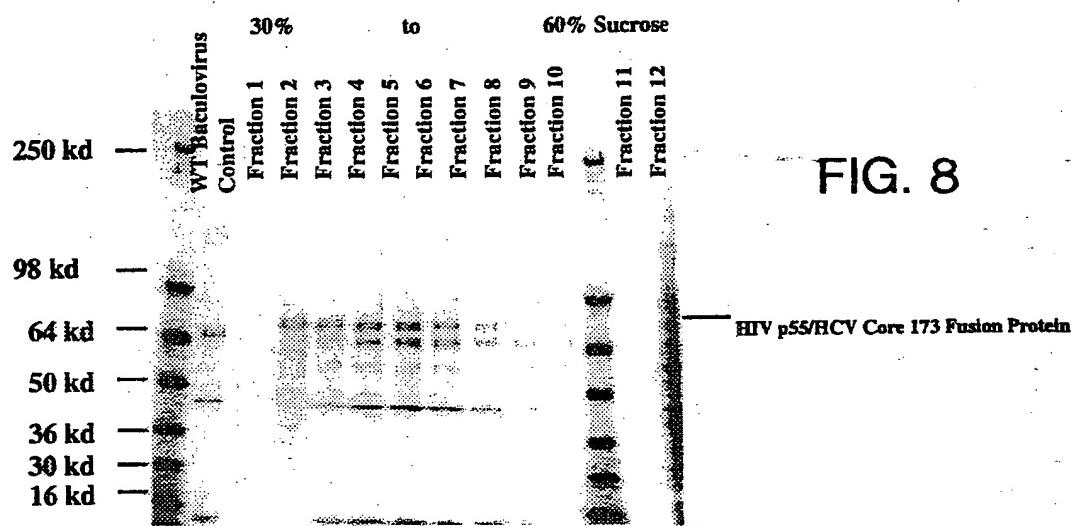


FIG. 8

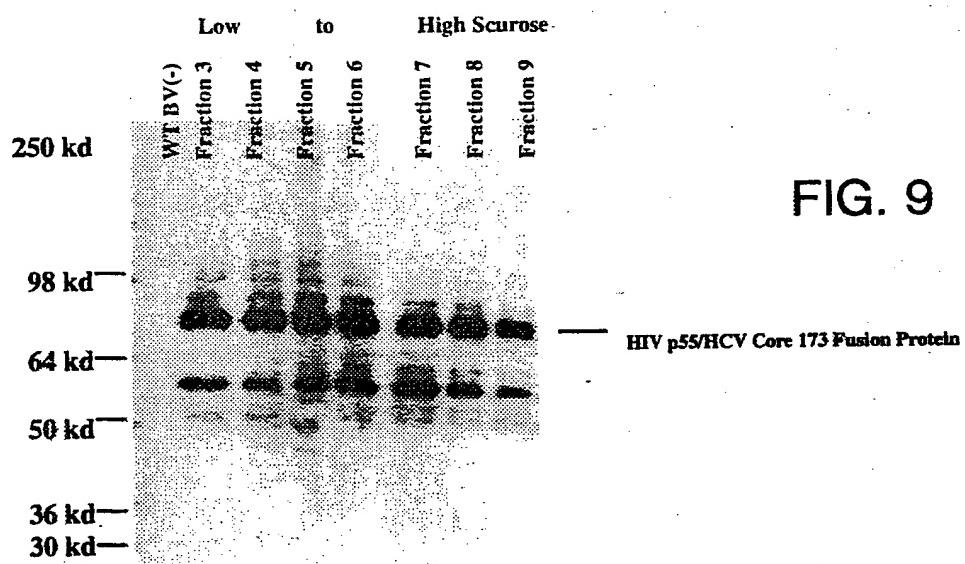
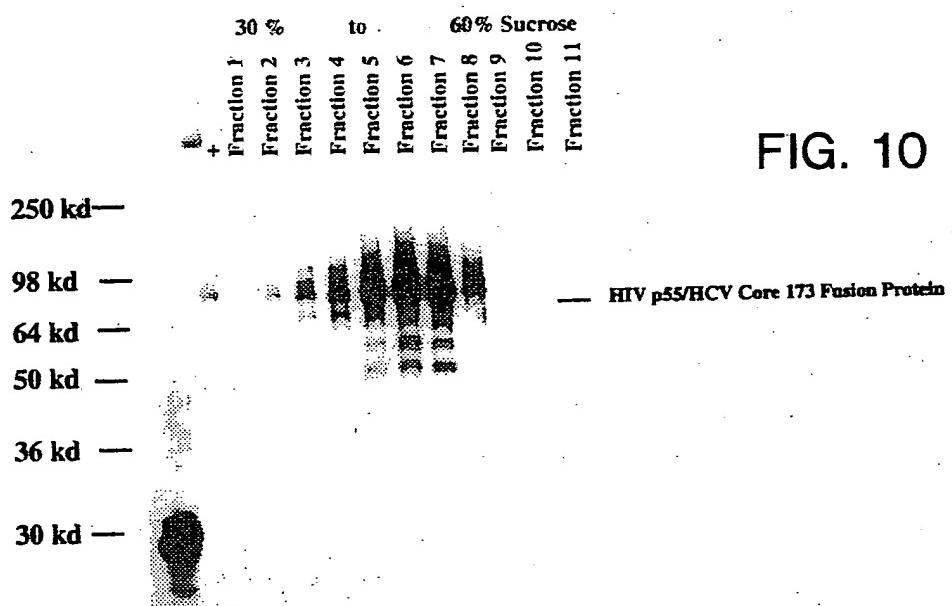


FIG. 9

17 / 131



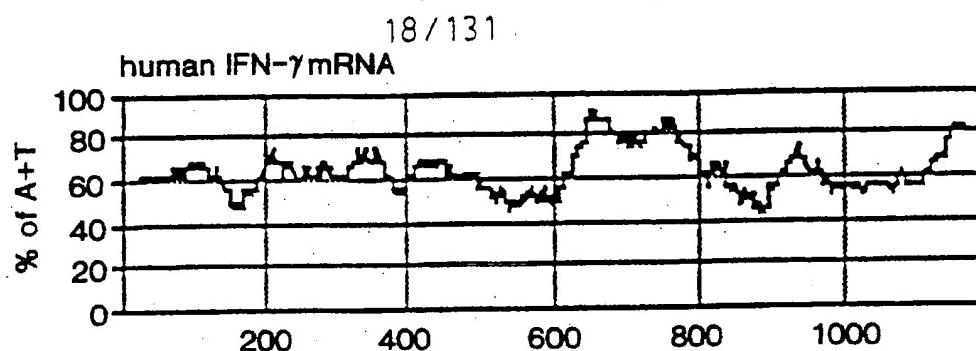


FIG. 11A

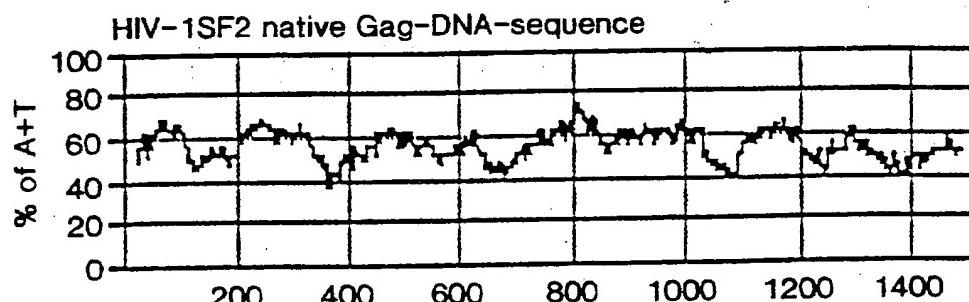


FIG. 11B

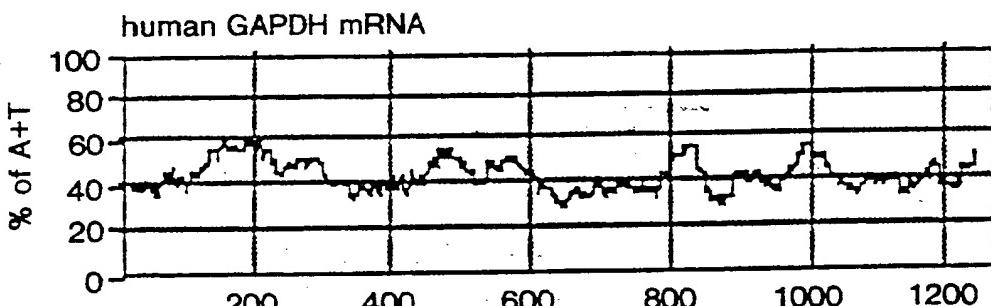


FIG. 11C

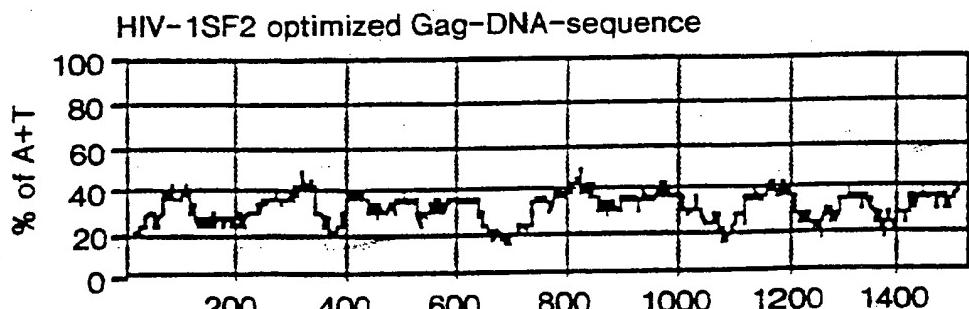


FIG. 11D

19 / 131

FIG. 12A

20 / 131

CAGGGATGTAAAATTGGTGCAGAAAACCTTGTGGTCAAAATGGAAACCCGAGTTGTAAGCATTTAAGGCA
 Inact. 7 TTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAAGGGAGTGGGGACCCGGCATAAAGCAAGAGTT
 C C C G T

Inact. 8 TTGGGTGAAGCCATGAGCCAAAGTAACATAATCAGCTAACATAATGATGGAGAGGGCAATTTAGGAACCAAAGAAAAG
 C G G G C G G C
 Inact. 9 ACTGTTAAGGGTTTCATTTGTGGCAAAAGAAGGGCACATAGCCAAAAAATTGGGGCCCTAGGGAAAAGGGCTTTGG
 C C G G C C C C G

AGATGTGGAAAGGAAACCAAAATGAAAGATTGCACTGAGAGACAGGCTAATTAGGGAAAGATCTGGCCTTC
 TACAAGGGAAAGGCCAGGGAAATTTCAGAGCAGACCAGAGGCCAACAGCCCCAACAGAGGAGCTTCAGGGTTGG
 GAGGAGAAACAACTCCCTCTCAGAAGCAGGGCGGATAGACAAGGAACGTATCCTTTAACCTCCCTAGATCACT
 Inact. 11 TTTGGCAACGACCCCTGGTCACAAATAAGGATAAGGGCAACTAAAGGAAGCTCTATTAGATACAGGGAGATGATA
 G C C G GC G
 CAGTATTAGAGAAATGAAATTGGCAGGGAAATAATGGAAACCAAAATGATAAGGGGAATTGGAGTTTATCAAAAGTA
 Inact. 12 GACAGTACGATCAGATACCTGTAGANATCTGGACATTAAGCTATAGGTACAGTATTAGTAGGCCTAACCTGTCA
 G
 ACATAATTGGAAAGAAATCTGGTGAATCAGATGGITGACTTTAAATTTCCCCATTAGTCCTATTGAAACCTGACAG
 Inact. 13 C C C C G C C C
 TAAATTAAAGCCAGGAATGGATGGCCAAAAGTTAACCAATGGCATTGACAGAAGAAAAATAAAAGCATTAGTAG
 Inact. 14 G G G C C G C

FIG. 12B

21 / 131

FIG. 12C

AGATATGTACAGGAAATGGAAAAGGAGGGAAAAATTCTAAAATTGGGCCCTGAAAATCCATAAAACTACCAACGACTCT
CTTAAAGAAAAAGACAGTACTAATGGAAAAGACTAGTAGATTCTAGAGACTTAATAAAAGAACTCAAGACTAAC
GGGAAGTTCAGGTAGGAAATACCACACCCCGAGGGTAAAAGAAAATCAGTNACGTTAGGATGTGGGTGATG
CATACTTTTCAGTTCCCTAGATAAGACTTAAAGGTTAGAAGTATCTGATTACCTAGTACGTTACCATAC
CAGGGATTAGATAATCTGACAATGGTGTGCCACAGGGATGGAAAAGGATCAGGCAAAAGCTTACAGGATGAGGAA
AAATCTTAGGGCTTTAGAAAACGAAATCCAGACATAGTTATCTTAACTACATGGATGATTGTGAGGATCTG
ACTTAGAAAATAGGGCAGCATAGAAACAAAATAGGAAACTGGAGACAGCATCTGTTAGAAGTATCTG
ACAAAACATCAGAAAAGAACCTCCATTCTTGGATGGGTTAGRACTCCATCTGATAAAATGGACAGTACAGCTTA
TRATGCTGCCAGAAAAGACAGGCTGACTGTCAATGACATACAGAGTTAGTGGAAAATTTGAGTGGGAA
TTTATGCGGGGATTAAAGTAAAGCAGTTATGTAACCTCTTAACTACAGGAACTAACAGGAA
CAGGAAAGCAGGAGCTAGAAACTGGCAGAAAACAGGGAGATCTAAAGAACCTGCAATATGGACAT
CAAAGACTTATGAGCAAATACAGAACCTCCATTCTTGGATGGGTTAGRACTCCATCTG
ATCTGAAAACAGGAAGTAGTATGCAAGGATGAGGGTGCCTACACTAATGATGTTAAACAGTAACTCC
ANGSTATCCACAGAAAGCATAGTAAATGGGAAAGATTCTTAATTTAACTCCATTCTGAGTGGGAGTT
CATGGTGGATGGAGTATGGCAAGCTACCTGGATTCTGAGTGGGAGTT
GGTACCAAGTTAGAAAAGAACCCATAGTAGGAGCAGAAAACCTCTATG
TAGGAAAAGCAGGATATGTTACTGACAGGAAAGACAAAAGTGTCTCCATTAGCTGACACACAAATCAGAGACT
AATTACAGCAATTCACTCTAGGTTGAGGATTGGGATTAGAAGTAAACATGTAACAGCTCACAATGCTAA
GAATCATTCAGCACAAACAGATAAGGTGAATCAGGTTAGCTCAGTCAATCTGAGGAAATTAATAGGCA
AGGTCTACCTGGCATGGTACAGGACACAAAGGAATTGGGAAATTTAGGAGTT
TCAGGAAAGTACTATTGTAATGGGAAATAGATAAGGCAAGGAACTGAAATACATAGCA
TGGCTAGTGTGTTAACCTGCCACCTGTAGTAGCAAAAGGAATTAGTAGCCAGCTGTGATTAAGGAA
AAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGTGA
TGTTAGCAATTCTGAGCTGCACTGGGATGATATAAGGCAAGAAGTT
TTCTCTTAAATTAGCAGGAAAGATGGCCAGTAAAAACAAATACATAGCA
TTAAGGCCCTGTTGGCAGGGTCAAGCAGGAATTGCA
AAATCTATGAATAATGAAATTAGGACAGGTTAGAAGGAAATAGTAGAGCATA
TGGCAGTATTCACTCCACAAATTAAAGAAAAGGGGGATTGGGGATACAGTGCAGGG
TTCTCTTAAATTAGCAGGAAAGATGGCCAGTAAAAACAAATACATAGCA
TTAAGGCCCTGTTGGCAGGGTCAAGCAGGAATTGCA
AAATCTATGAATAATGAAATTAGGACAGGTTAGAAGGAAATAGTAGAGCATA
TGGCAGTATTCACTCCACAAATTAAAGAAAAGGGGGATTGGGGATACAGTGCAGGG
TTAGGAAAGGACTACAAAGCAAAATTACAAAAGGCAAGCTCTGGAAAAGGTGAGGGCAGTAA
ACAAAGATCCCCCTTGGAAAGGACAGCAAGCTCTGGAAAAGGTGAGGGCAGTAA
ACATAAAAGTAGTGTGCCAGAAAGCAAAATTCACTCCACAAATTAAAGAAA
CAGTAGCAGGATGAGGATTAG

22 / 131

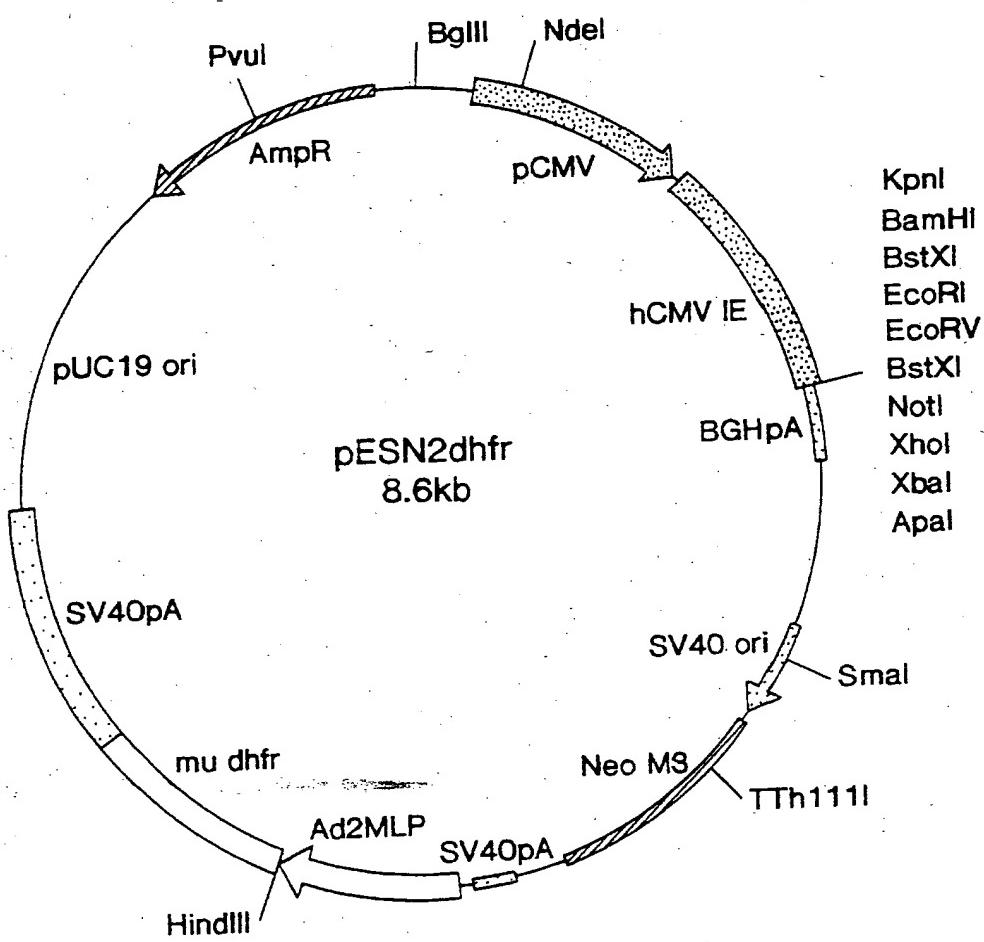


FIG. 13A

23 / 131

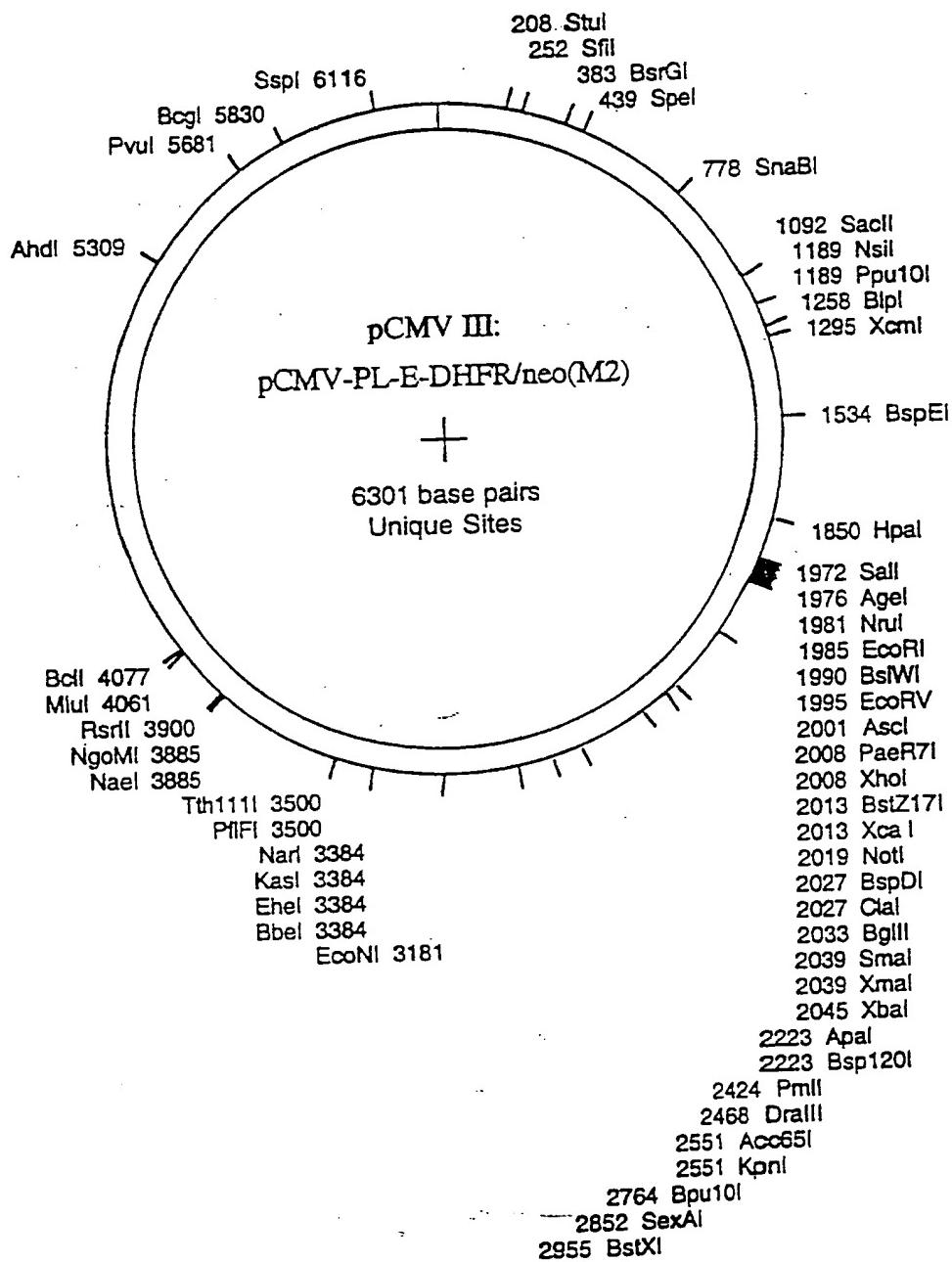


FIG. 13B

24 / 131

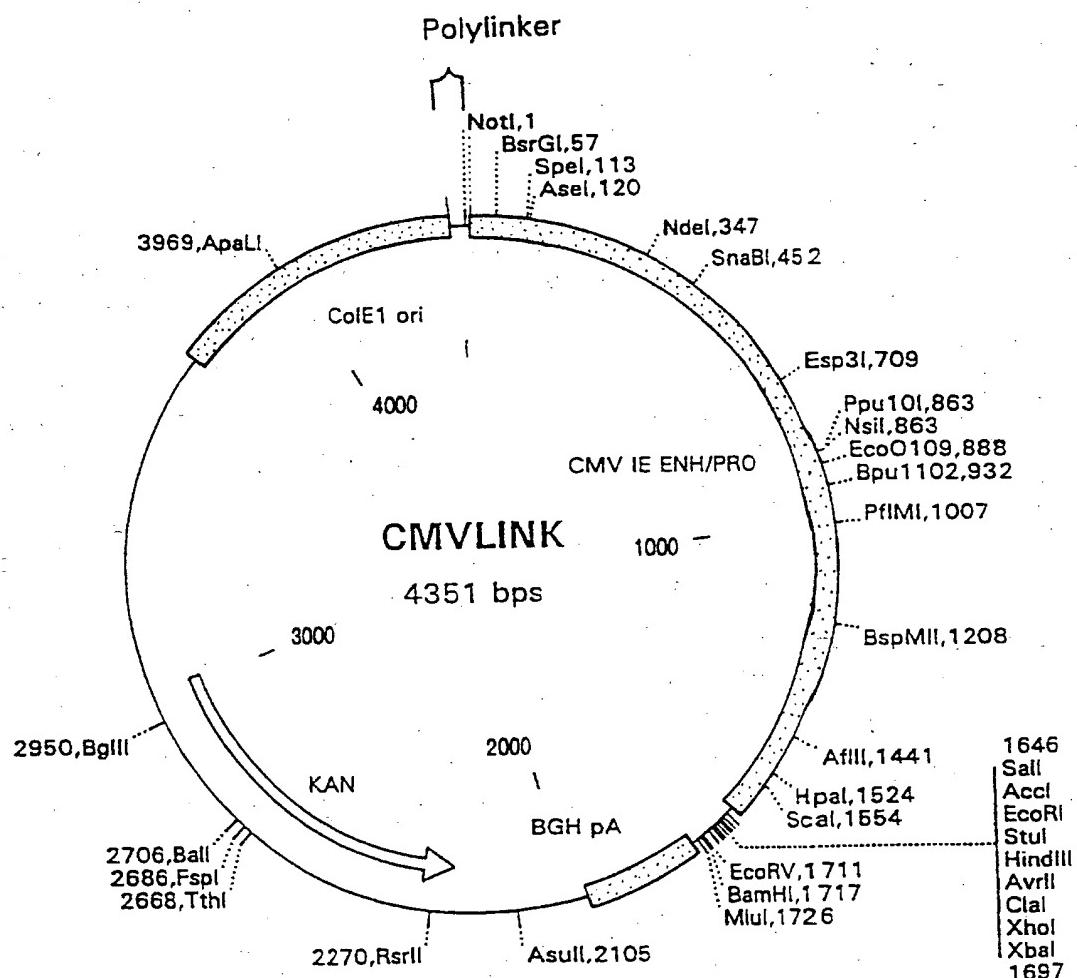


FIG. 14

25 / 131

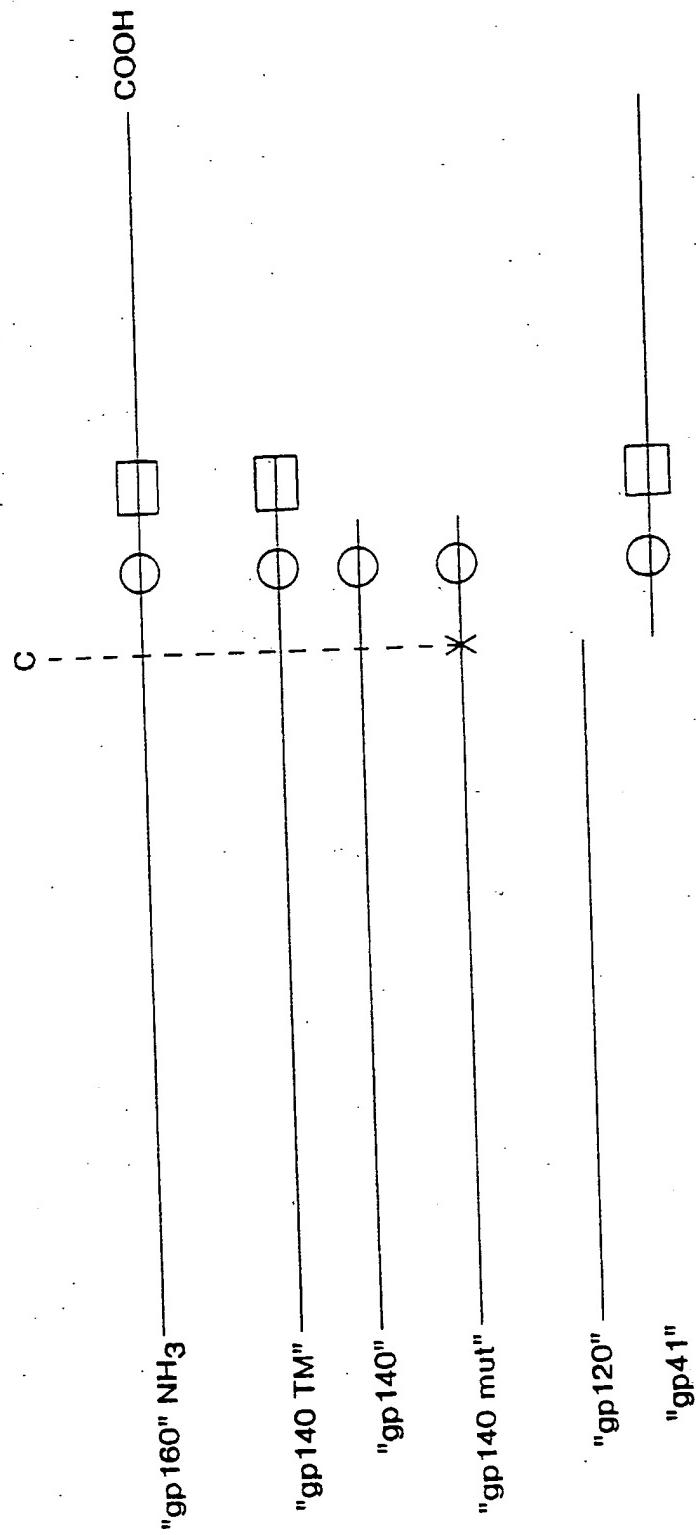


FIG. 15

26 / 131

9P120wtsf162

GTAGAAAAATTGTGGTCAACAGTCTATTATGGGTACACAGTCAAAAGGTTACATAATGCTGGCCACACATGCCCTGTGTACCCAC
 AGACCCCTAACCCACAAGAAATTAGTATTGGAAAATGGACAGAAAATTTCACATGTGGAAAAAATTAACATGTGGAAAGTTAACATG
 GTAGAACAGATGCAAGGATATAATCAGTTATGGTCAAAAGTCTAACGGTCAAGGATCTAACAGTCAAGGAAAGTTAACCC
 CACTCTGTGTFACTCTACATTGCAACTAAATTGAAGGAATTGCTCTCTTCAGGTCAACCACAGCATAAAGAATAAGATGCAGAAA
 GATGGACAGAGGGAAATAAAAATTGGCAACTAACTGTGTTATAAACTTGATGTAGTACCAATAGATAATAACAGCTAAATTGATAA
 GAATATGCACATTTCATTAACCTTGATGTAGTACCAATAGATAATAACAGCTAAATTGATAATAACAGCTAAATTGATAA
 ATTGTAAACACCTCAGTCATTACACAGGCCGTGTCCAAGGTATCCTTGAACCAATTCCCATACATATTG
 TCCCCCGGCTGGTTTGCATCTAAAGTGTAAATGATAAGAAGGTCAATGGACATGGACATGTACAAT
 GTCAAGCACAGTACAATTGTACACATGGAATTAGGCCAGTAGTGTCAACTCAATTGCTTAATGGCAGTC
 TAGCAGAAAGGGGTAGTAATTAGATCTGAAATTTCACAGACAATTGCTAAATTGATAATAGTACAGCT
 GAGGAAATCTGTAGAAATTAAATTGTACAAGACCTAACAAATAACAGAAAAGTATAACCTATAGGCCG
 GGGAGAGCATTATGCAACAGGGAGACATAATGGAGATAAGACAAAGCACATTGTAACATTAGTGGAG
 AAAAATGGAAATAACACTTTAAACAGATACTTACAAAATACAGATACTTACAAAATACAGCAAAATTGGAAATAACAAATAGT
 CTTTAAGCAATCCTCAGGAGGGACCCAGAAATTGTAATGCAAGTTTAATTTGAGGGGAAATTTC
 TACTGTAAATTCAACACAGCTTTTAATAGTACTTGGAAATAACTATAGGCCAAATAACACTAATGGAA
 CTATCACACTCCCATTGCAAAATAACAAATTAAACAGGTGGCAGGAAGTAGGAAANGCAATGTATGC
 CCCTCCCATCAGGGACAAATTAGATGCTCATCAAATTACAGGACTGCTATTAAACAGAGATGGTGGT
 AAAGAGATCAGTAACACCAGACATCTTCAGACCTGGAGGATATGGGAGACATTGGAGAAGT
 AATTATAATAATAAAGTAGTAAATTGAGCTTACAGGACACCCACCAAGGCAAAAGAGAAGAGT
 GGTCAGAGGAGAAAAAGA

FIG. 16
 (SEQ ID NO:30)

GTAGAAATAATTGTGGGTCACAGCTTATTATGGGGTACCTGTGGAAAGGAAGCACCACACTCTATT
GTGCATCAGATGCTAACGCCATGACACAGGGTACATAATTGTCTGGGCCACACATGCCCTGTACCCAC
AGACCCCTAACCCACAGAAATAATGTTGAAATGTGACAGAAANTTTAACATGTGGAAAATAACATG
GTAGAACAGATGCTAGGGATAATAATCAGTTATCGGATCAAAGCTTAAGGCTAAGGTAAAGTTAAC
CACTCTGTGTTACTCTACATTGCACTAAATTGCTTCAAGGTCAACCAAGCATAAGAAATAAGATGGAGAA
GATGGACAGAGGGAAATAAAAATTGGCTTCAAGGTCAACCAAGCATAAGAAATAAGATGGAGAA
GAATATGCACTTTTATAACCTTGATGTTAGTACCAATAAGATAATGATAATAGATAATAGATA
ATTGTAACACCTCAGTCATTACACAGGCCCTGTCCAAGGTATCCTTGAACCAATTCCCATACATTATTG
TGCCCCGGCTGGTTTGCGATTCTAAAGGTAAATGATAAGAAGTTCAATGGATCAGGACCATGTAACAAAT
GTCAGGCACAGTACAATGTACACATGGAAATTAGGCACAGTAGTGTCAACTCAATTGCTGTTAAATGGCAGTC
TAGCAGAAAGAAGGGTAGTAATTAGATCTGAAAAATTTCACAGACAATGCTAAACACTATAATAGTACAGCT
GAAGGAATCTGTAGAAATAATTGTACAGACCTAACATAAGAAGATAATAAGACAAAGCACATTGTAACATTAGTGGAG
GGGAGAGGCATTATGCAACAGGAGACATAATAGGAGATAATAAGACAAAGCACATTGTAACATTAGTGGAG
AAAAAATGGAATAACACTTAAACAGATAAGTTACAAAATTACAAGCAAAATTGGGAATAACAAATAGT
CTTAAGGCAATCCTCAGGGGACCCAGAAATTGTAATGCAAGCTTAAATTGTTAATTTGAGGGGAATTTTTC
TACTGTAATTCAACACAGCTTTAATAGTACTTGGAAATAACTATAGGACTCTGCTATAACAGAGATGGTGGT
CTATCACACTCCCATTGCAAAATTAGTCATCAAATACTAGGACTCTGCTATAACAGTGGGAAGTAGGCAATGTT
CCCTCCCATCAGGGACAATTAGTCATCAAATACTAGGACTCTGCTATAACAGTGGGAAGTAGGCAATGTT
AMAGAGATCAGTAACACCCAGAGATCTCAGACCTGGAGGTGGAGATAATGGGACAAATTGGAGAAAGTGA
AATTATAATAATAAGTAGTAAATTAGCCATTAGGAGTAGGCCACCAAGGCAAAAGGAAGAGT
GGTGCAGAGAGAAAAAGAGCAAGTGAAGCTAGGAGCTAGGAGCTAGGAGCTATGTTCTGGGTCTGGGTG
ACTATGGGGCACGGTCACTGACGGCTGACGGTACAGGCCAGACAATTATGCTGGTATACTGCAACAGC
AGRACAAATTGCTGAGAGGCTATTGAGGGCAACAGGCAACTCTGTTGCAACTCACAGTCTGGGCATCAAAGCA
GCTCCAGGCAAGAGTCCTGGCTGGAAAGATAACCTAAAGGATCAACAGCTCCCTAGGGATTGGGTG
TCTGGAAAAACTCATTGCACCAACTGCTGGCTTGGAAATGCTAGTTGGAGTAATAATCTGGATCAGA
TTTGGAAATAACATGACCTGGATGGAGTGGAGAATAATGACAATTACACAAACTTAATAACACCTT
AATTGAAGGAATCGCAGAACCAACAGAAATAAGAATGAAACAGAAATTATTAGAAATTGGATAAGTGGCAAGT
TTGTGGAAATTGGTTGACATATCAAAATTGGCTGGGTATATA

FIG. 17

(SEQ ID NO:31)

28 / 131

gp160wtsf162

GTAGAAAAATTGGGGTACAGTCTATTATGGGTACCTGTGGAAAGAACCAACCACCTCTATTG
GTGCATCAGATGCTAAGCTATGACACAGAGGTACATAATGCTGGGCCACATGCCTGTGACCCAC
AGACCCCTAACCCACAAGAAATAGTATTGGAAAATGTGACAGAAAATTAACTGTGGAAAATAACATG
GTAGAACAGATGCATGAGGATATAATCAGTTATGGATCAAAGCTAAAGCCATGTGAAAGTTAACCC
CACTCTGTGTTACTCTACATTGACTAATTGAAGAATGCTACTAATACCAAGAGTAGTAATTGGAAAGA
GATGGACAGAGGAGAAAATTTGCTCTTCAGGTACCCACAAGCATAAGAAATAAGATGCAGAAA
GAATATGCACTTTTATAAACTTGATGTAGTACCAATAGATAATGATAATACAAGCTATAAATTGATAA
ATTGTAACACCTCAGTCATTACACAGGCTGTCCAAGGTATCCTTGAACCAATTCCCACATATTG
TGCCTCGGCTGGTTTGCATTCTAAAGTGTAAATGATAAGAAGTCAATGGATCAGGACCATGTACAAAT
GTCAGCACAGTACAATGTACACATGAAATTAGGCCAGTAGTGTCAACTCAATTGCTGTTAATGGCAGTC
TAGCAGAAGAAGGGTAGTAATTAGATCTGAAAATTACAGACAATGCTAAAACATAATAGTACAGCT
GAAGGAATCTGTAGAAATTAAATTGTACAGACCTAACATAATACAAGAAAAGTATAACTATAGGACCG
GGGAGAGCATTTATGCAACAGGAGACATAATAGGAGATATAAGACAAGCACATTGTAACATTAGTGGAG
AAAATGGAATAACACTTTAAACAGATAGTTACAAAATTACAAGCACAATTGGGAAATAAACAAATAGT
CTTTAAGCAATCCTCAGGAGGGACCCAGAAATTGTAATGCACAGTTTAATTGTGGAGGGGAATTTTC
TACTGTAATTCAACACAGCTTTAATAGTACTTGAATAATACTATAGGCCAAATAACACTAATGGAA
CTATCACACTCCCATGCAGAATAAAACAAATTATAACAGTGGCAGGAAGTAGGAAAAGCAATGTATGC
CCCTCCCATCAGAGGACAAATTAGATGCTCATCAAATTACAGGACTGCTTAAACAAGAGATGGTGGT
AAAGAGATCAGTAACACCACCGAGATCTCAGACCTGGAGGTGGAGATATGGGACAATTGGAGAAGTG
AATTATATAAATATAAAGTAGTAAAATTGAGCCATTAGGAGTAGCACCACCAAGGAAAGAGAAGAGT
GGTGCAGAGAGAAAAAGAGCAGTGCAGCTAGGAGCTATGTTCTTGGGTCTTGGGAGCAGCAGGAAGC
ACTATGGCGCACGGTCACTGACGCTGACGGTACAGGCCAGACAATTATTGCTGGTATAGTGCACACAGC
AGAACAAATTGCTGAGAGCTATTGAGGCGAACAGCATCTGCAACTCACAGTCTGGGCATCAAGCA
GCTCCAGGCAAGAGTCCCTGGCTGTGGAAAGATAACCTAAAGGATCAACAGCTCTAGGGATTGGGTTGC
TCTGGAAAACTCATTGACCACTGCTGTGCCCTTGGAAATGCTAGTGGAGTAATAATCTCTGGATCAGA
TTTGGAAATAACATGACCTGGATGGAGTGGGAGAGAGAAATTGACAATTACACAAACTTAATATAACCTT
AATTGAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGAATTATTGAAATTGGATAAGTGGCAAGT
TTGTGGAATTGGTTGACATACAAATGGCTGTGGTATATAAAATTCTATAATGATAGTAGGAGGTT
TAGTAGGTTAACGGATAGTTTACTGTGCTTCTATAGTGAATAGAGTTAGGCAGGGATACTCACCATT
ATCATTTCAGACCCGCTTCCCAGCCCCAGGGGACCCGACAGGCCAGAGGAATCGAAGAAGAGGTGGA
GAGAGAGACAGAGACAGATCCAGTCATTAGTCATGGATTATTGCACTCATCTGGGACGATCTACGGA
GCCTGTGCCTCTCAGCTACCAACCGCTGTGAGAGACTTAATCTGATTGCAAGCGAGGATTGTGGAACTTCT
GGGACGCAGGGGGTGGGAAGCCCTCAAGTATTGGGGAAATCTCTGCAGTATTGGATTCAAGAACAAAG
AATAGTGTGTTAGTTGTTGATGCCATAGCTATAGCAGTAGCTGAGGGGACAGATAGGATTATAGAAG
TAGCACAAAGAACCTGGTAGAGCTTCTCCACATACCTAGAAGAATAAGACAGGGCTTGAAAGGGCTT
GCTATAA

FIG. 18

(SEQ ID NO:32)

gp120.modSF162

FIG. 19
(SEQ ID NO:33)

30 / 131

FIG. 20
(SEQ ID NO:34)

gp120.modsF162.delV1v2

FIG. 21
(SEQ ID NO:35)

SUBSTITUTE SHEET (RULE 26)

32 / 131

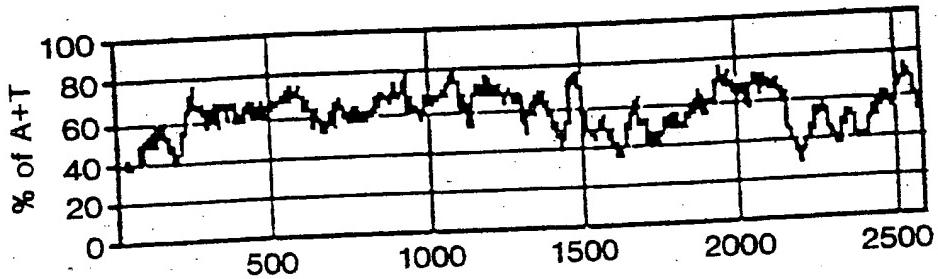


FIG. 22A

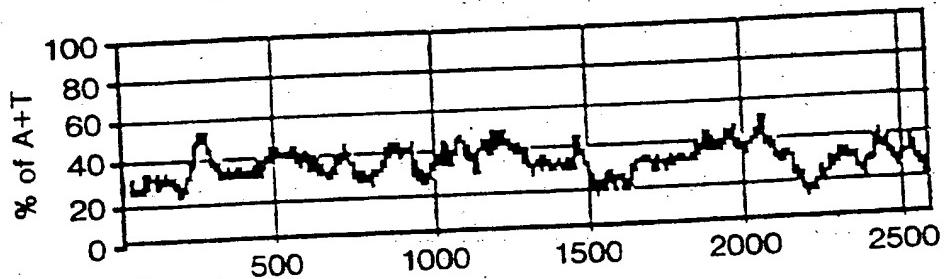


FIG. 22B

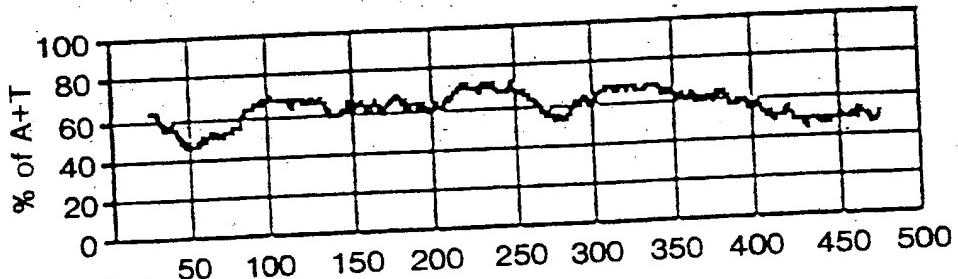


FIG. 22C

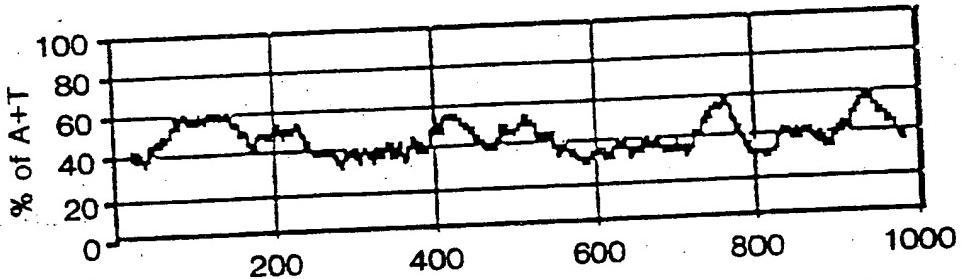


FIG. 22D

33 / 131

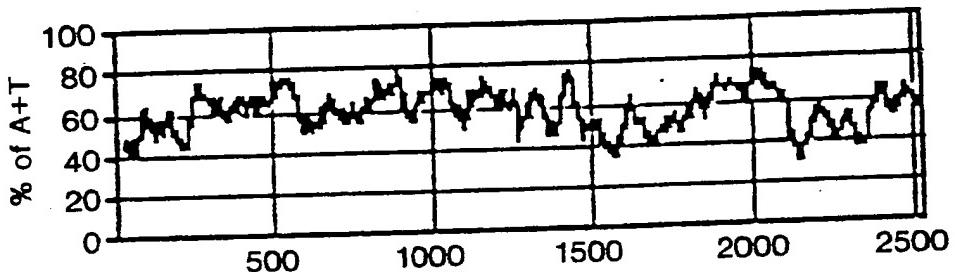


FIG. 22E

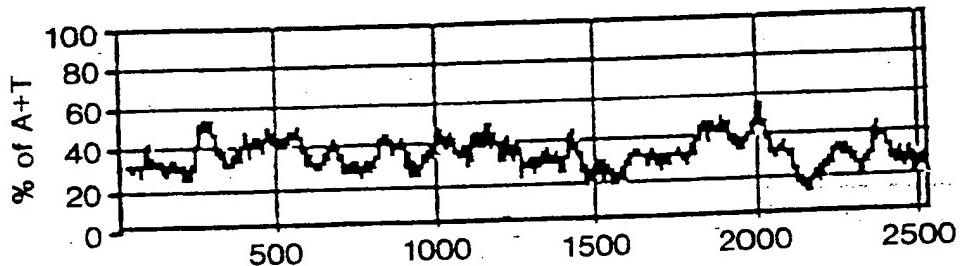


FIG. 22F

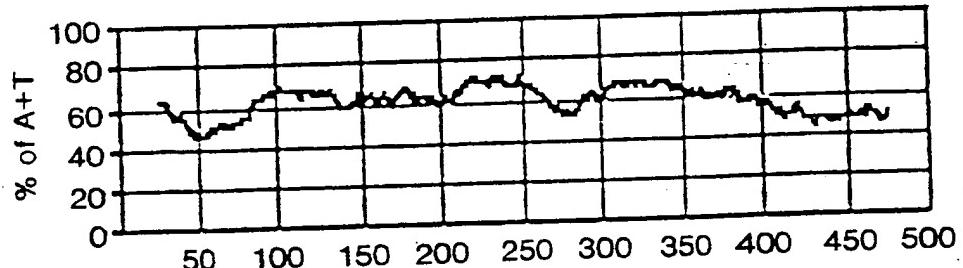


FIG. 22G

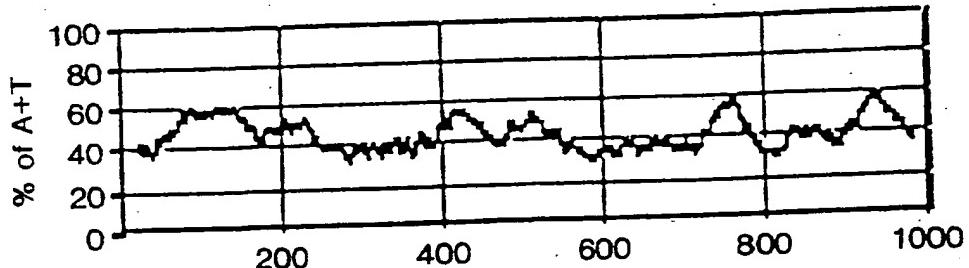


FIG. 22H

34 / 131

gp140.modSF162

gaattcgccaccatggatgcaatgaagagagggctctgtgtgtgtgtgtggaggcagtc
ttcggttcgccccagcgccgtggagaagactgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgccagcgacgccaaggcctacgacaccgagggtgcacaacgtg
tggccaccacccacgcctgcgtgcccaccgaccccaaccccaaggagatcgtgtggagaacacgtgacc
gagaacttcaacatgtggaaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacacctg
aagaacgccaccaacaccaagagcagcaactgaaaggagatggaccgcggcgagatcaagaactgc
agcttcaaggtgaccaccagcatccgcaacaagatgcagaaggagtacgcccctgttctacaagctg
gacgtggtgcccatcgacaacacaccagactacaagctgatcaactgcaacaccagcgtgatc
acccaggcctgccccaaagggtgagcttgcagccccatccccatccactactgcgcgcggccggcttc
gccccatccctgaagtgcaacgcacaagaagttcaacggcagcggccctgcaccaacgtgagcaccgtg
cagtgcacccacggcatccgccccgtggtagcaccagctgctgtgaacggcagcctggccgag
gagggcgtggtagccgcagcagaacttcaccgacaacgccaagaccatcatcgtgcagctgaag
gagagcgtggagatcaactgcacccggccacaacaacaccccgcaagagcatcaccatcgcccc
ggccgcgccttctacgccaccggcagacatcatcgccgacatccgcaggcccactgcaacatcagc
ggcgagaagtggaaacaacaccctgaagcagatcgtgaccaagctgcaggcccagttcggcaacaag
accatcgtttcaaggcagcagcggcggcggcggcggcggcggcggcggcggcggcggcggc
ggcgagttttactgcacacagcaccctgatgttcaacacagcacctggaaacaacaccatcgcccc
aacaacaccaacggcaccatccccatccctggccatcaaggcagatcatcaaccgtggcaggag
gtggcaaggccatgtacccccccatccgcggccagatccgctgcagcagcaacatcaccggc
ctgctgcgtgacccggcggcggcggcggcggcggcggcggcggcggcggcggc
ggcgacatgcgcgacaactggcgcagcggcggcggcggcggcggcggcggc
ggcgtggccccccaccaaggccaagcggcggcgtggtagcggcggcggcggcggc
ggcgtggccccccaccaaggccaagcggcggcgtggtagcggcggcggcggcggc
gtggcggcgtacccatccgcggccagatccgcggccagatccgcggccagatccgc
gtggcggcgtacccatccgcggccagatccgcggccagatccgcggccagatccgc
gtggcggcgtacccatccgcggccagatccgcggccagatccgcggccagatccgc
gtggcggcgtacccatccgcggccagatccgcggccagatccgcggccagatccgc
accaccggcgtggccatccgcggccagatccgcggccagatccgcggccagatccgc
acccggatggagttggagcggcggcggcggcggcggcggcggcggcggcggc
agccagaaccacggcaggagaagaacacggcggcggcggcggcggcggcggc
aactgggttcgacatcagcaagtggctgtggtagatctaactcggc

FIG. 23
(SEQ ID NO:36)

35 / 131

gp140.modsF162.delV2

gaattccaccatggatcaatgaagagagggctctgtgtgctgctgtgtggaggcagtc
ttcgtttcgcccagcgcgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgccagcgcacggcatacgacacccgaggtgcacaacgtg
tggccaccacccacgcctgcgtgccaccgaccccaaccccaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgttggagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacctg
aagaacgcaccaacaccaagagcagcaactggaaggagatggaccgcggcgagatcaagaactgc
agcttcaaggtgggcgcggcaagctgtatcaactgcaacaccagcgatcaccaggcctgcccc
aaggtgagcttcgagccatccccatccactactgcgcggccgcggcgttcgcacatcctgaagtgc
aacgacaagaagtcaacggcagcggccctgcaccaacgtgagcaccgtgcagtgcacccacggc
atccgcggccgtggtgagcacccagctgtgtgaacggcagcctggccgagggaggcgtgggtgatc
cgcagcagaacttccaccgacaacgccaagaccatcatcgtgcagctgaaggagagcgtggagatc
aactgcacccggcccaacaacaacacccgcaagagcatcaccatcgcccccggccgccttctac
gccaccggcgacatcatcgccgacatccggcaggcccactgcaacatcagcggcgagaagtggAAC
aacaccctgaagcagatcgtgaccaagctgcaggcccagttcggcaacaagaccatcgtgttcaag
cagagcagcggcgccgaccccgagatcgtgtatgcacagcttcaactgcggcggcgagttcttctac
tgcaacagcaccctggttcaacagcacttggaaacaacaccatcgcccccacaacaccaacggc
accatcaccctgcgcgtccatcaagcagatcatcaaccgcgtggcaggaggtggcaaggccatg
tacgccccccatccggccgcagatccgcgtgcagcagcaacatcaccggccgtgtgcacccgc
gacggcggcaaggagatcagcaacaccacccgagatcttccggccggccgacatgcgcgac
aactggcgcagcggagctgtatacaaggtggtaagatcggagccctggcgtggccccacc
aaggccaaggcgcgcgtggcagcgcgagaagcgcgcgtgacccctggcgcgcgtgttccctggc
tccctggcgcgcgcggcagcaccatggcgcggcagccctggcgcgtgcaggcccgcag
ctgcgtgagcggcatcgtgcagcagcagaacaacctgtgcgcgcgcgtggcggccaccctg
ctgcagctgaccgtgtggcatcaagcagatcgtgcaggcccgcgtgtggccgtggagcgtaccc
aaggaccagcagctgtggcatctgggctgcagcggcaagctgtatctgcaccaccgcgtgccc
tggaaacgcgcagctggagcaacaagagcctggaccagatctggaaacaacatgacccctggatggagtg
gagcgcgagatcgcacaactacaccaacctgtatctacaccctgtgcaggagagccagaaccagcag
gagaagaacgcgaggagctgtggagctggacaagtggccagccctggactggatggatc
agcaagtggctgtgttacatctaactcgag

FIG. 24
(SEQ ID NO:37)

36 / 131

gp140.modSF162.delV1V2

FIG. 25
(SEQ ID NO:38)

37 / 131

gp140.mut.modsF162

gaattcgccaccatggatcaatgaagagagggctctgcgtgtgctgctgtgtggagcagtc
ttcgccccccaggccgtggagaagactgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
gagggcaccaccacccctgttctgcgccagcgcacgccaaggcctacgacaccgaggtgcacaacgtg
tgccacccacgcctgcgtgcccaccgaccctaaccggagatcgtgctggagaacgtgacc
gagaacttcaacatgtggaaagaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacctg
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcggcgagatcaagaactgc
agcttcagggtgaccaccaggcatccgcaacaagatgcagaaggagtacgcctgttctacaagctg
gacgtggtgcccatcgacaacacaccaggatacaagctgatcaactgcaacacaccaggctgatc
acccaggcctgcccccaaggtagctcgagccatccccatccactactgcgcgcgcgcggcttc
gccatcctgaagtgcacgcacaagaagttcaacggcagcggccctgcaccaacgtgagcaccgtg
cagtgcacccacggcatccgccccgtggtagccacccagctgcgtgacgcgcctggccgag
gagggcgtggtagccgcagcggagaacitcaccgcacaacgccaagaccatcatcgtcagctgaag
gagagcgtggagatcaactgcacccgcaccaacaacacaccgcacagagcatcaccatcgcccc
ggccgcgccttctacgcacccggcgcacatcatcgccgacatccgcacggccactgcaacatcagc
ggcgagaagtggaaacaacacccctgaagcagatcgtgacccatgcggccagttoggcaacaag
accatcgtgttcaagcagagcagcggcggcgcaccccgagatcgtgatgcacagctcaactgcggc
ggcgcaggcttctactgcaacagcacccagctgttcaacagcacctggaaacaacaccatcgcccc
aacaacaccaacggcaccatcacccctgcccgcacatcaagcagatcatcaaccgcgtggcaggag
gtggcaaggccatgtacgccccccatccgcggccagatccgctgcagcagcaacatcaccggc
ctgcgtgacccgcgcacggcggcaaggagatcagcaacaccaccgagatctccgcgcgcggc
ggogacatgcgcgacaactggcgacgcgagctgtacaaggatcaagggtggtagagatcgagccctg
ggcgtrggccccccaccaaggccaagcgcgcgtggcgacgcgcgcgagaagagcgcgcgtgaccctggc
gccatgttcttggcttctgggcccgcggcgcgcacccatgggcgcgcgcgcgcgcgcgcgc
gtgcaggccccgcgcacccgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
gcccagcagcaccctgcgcagctgacccgtgtgggcatacgacgcgcgcgcgcgcgcgc
gtggagcgcacccgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
accaccgcgcgtgcctggaaacgcgcacgcgcgcgcgcgcgcgcgcgcgcgc
acctggatggatggggagcgcgcagatcgacacactacaccaacctgatctacaccctgatcgaggag
agccagaaccagcaggagaagaacgcgcaggagctgtggagctggacaagtgccgcgcgcgc
aactggatcgacatcagcaagtggctgtggatcatctactcgag

FIG. 26

(SEQ ID NO:39)

38 / 131

gp140.mut.modsF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgtgtgtgtgtgtggaggcagtc
ttcgttcgcccacgcgcgtggagaagctgtgggtgaccgtgtactacggcgtccccgtgtggaaag
gaggcaccaccacccctgttctgcgccagcgcacgccaaggctacgacaccgagggtgcacaacgtg
tggccacccacgcctgcgtgccaccgaccccaaccccaaccccaaccccaaccccaaccccaaccc
gagaacttcaacatgtggaaagaacaacatggtgaggcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacctg
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcggcgagatcaagaactgc
agcttcaaggtgggcggcggcaagctgtatcaactgcaacaccagcgtgatcacccaggcctgcccc
aaggtgagcttcgagcccatccccatccactaictgcgccccccggccggcttcgccccatcctgaagtgc
aacgacaagaagttcaacggcagcggccctgcaccaacgtgagcaccgtgcagtgcacccacggc
atccggccctgtggtaggcacccagctgtgtgaacggcagcctggccgaggaggggcgtggtagtc
cgccagcggagaacttacccgacaaacgccaagaccatcatcgtgcagctgtgaaggagagcgtggagatc
aactgcacccgccccacaacaacacacccgcaagagcatcaccatcgccccccggccgcgccttctac
gccaccggcgacatcatcgccgacatccggcaggcccactgcaacatcagcggcgagaagtggaaac
aacaccctgaagcagatcgtgaccaagctgcaggcccagtctggcaacaagaccatcgtgtcaag
cagagcagcggcggcgaccccgagatcgtgatgcacagctcaactgcggcgccgagttcttctac
tgcaacagcaccagctgtcaacacgcacccctggaaacaacaccatcgcccccaacaacaccaacggc
accatcaccctggccatcgccatcaagcagatcatcaaccgcgtggcaggagggtggcaaggccatg
tacgccccccatccgcggccagatccgcgtgcagcacaatcaccggcctgtgtgcacccgc
gacggcggcaaggagatcagcaacaccaccgagaaccttccgccccccggccggccgacatgcgcgc
aactggcgagcagctgtacaagtgatcaagggtgtgaagatcgagccctggcgtggccccccacc
aaggccaagcgccgcgtggtagcgcgcgagaagagcgcgtgaccctggcgcacatgttccctggc
ttccctggcgccgcggcagcaccatggcgccgcgcgcgtgaccctgtgcaccgtgcaggccgc
ctgctgagcggcatcgtgcagcagcagaacaacctgtgcgcgcgcacatcgaggcccgacaccc
ctgcagctgacccgtgtggcatcaagcagctgcaggccgcgtgtggcgtggagcgtaccc
aaggaccagcagctgtggcatctgggtgtgcagcggcaagctgtatctgcaccaccgcgtgc
tggaaacgcgcagctggagcaacaagagcctggaccagatctggaaacaacatgaccctggatggagtg
gagcgcgagatcgcacaactacaccaacccgtatcaccctgtgcaggagagccagaaccaggc
gagaagaacgcgaggagctgtggagctggacaagtggccagccgtggactgggttcgacatc
agcaagtggctgtggatcatctaactcgag

FIG. 27

(SEQ ID NO:40)

39 / 131

gp140.mut.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggcttgctgtgtgtgtgtggaggcagtc
ttcgtttgcgccagcgccgtggagaagctgtgggtgactgttactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgcagcgcacgccaaggcctaagcacaccgagggtgcacaacgtg
tggggccaccacccacgcctgcgtgcccaccgaccccaacccccaggagatcgtgtggagaacgtgacc
gagaacttcaacatgttggaaagaaaaacatggggagcagatgcacgaggacatcateagcctgtgg
gaccagagccttgaaggccctgcgtgaagctgtgaccccccgtgcgtggggccggcaactgcccagacc
agcgtgatcacccagggcctgcccagggtgagcttcgagccatccccatccactactgcgcggcc
gccggcttcgcacatcttgcataacttgcacaaacggcagggggccctgcaccaacgtg
agcacccgtgcagtgcacccacggcatccggccctgtgtggagcaccacccagctgtgtgtgaacggcagc
ctggccgaggagggcgtggatccgcagcgagaaccttaccgacaacgcacagaccatcatcg
cagctgaaggagagcgtggagatcaacttgcacccgcaccaacaacaacacccgcaagagcatcacc
atcgcccccggccgccttctacgcacccggcgacatcatcgccgacatccggccaggcccactgc
aacatcagccggcgagaagttggaaacaacacccctgaagcagatcgtgaccaagctgcaggcccagg
ggcaacaagaccatcggttcaaggcagagcagccggcgaccccgagatcgtgatgcacagcttc
aactgcggccggcgagttcttacttgcacacgcaccccgactgttcaacagcacctggaaacaacacc
atcgccccaacaacacccaacccgcacccatcacccctgcctgcgcacatcaagcagatcatcaacccgc
tggcaggagggtggcaaggccatgtacgcggccatccgcggccagatccgcgtgcagcagcaac
atcaccggccctgcgtgtgacccgcacccgcggcaaggagatcagcaacaccaccgagatcttccgc
cccgccggccggcgacatcgccgacactggcccgaccccgactgtacaagggtggtaagatc
gagcccccctggccgtggcccccaccaaggccaagccgcgtggcagccgcgagaagagccgc
accctggccgccttgcgtggccatgttgcgtggccatccgcggccggcagccatggccgcggccgc
accctgaccgtgcaggcccgccagctgtgagccgcacatcggtgagccgcacccgc
cccatcgaggcccagccgcacccgtgtgcagccgtgtggccatccgcgtggccgc
gtgctggccgtggagccgttgcgtggccatccgcgtggccatccgcgtggccgc
ctgatctgcaccacccgcgtgcctggaaacgcgcagctggagcaacaaggccctggacc
aacaacatgacccatggatggagatggagccgcgcgacactacaccaacccgtatc
atcgaggagagccagaaccaggcaggagaagaacgcaggagctgcggagatggaca
gcctgtggaaactggatccgcacatcgcaactgggttgcgttatcatcttacttgcag

FIG. 28

(SEQ ID NO:41)

40/131

gp140.mut7.modsF162

gaattcgccaccatggatcaatgaagagagggctctgtgtgtgtgtgtgtggagcagtc
ttcgccccccaggccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgogccagcgcacgccaaggcctacgacacccgaggtgcacaacgtg
tggccaccacccacgcgtgcgtgcccaccgaccccaaccccaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggaaagaacaacatgtggagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgacccctgtgcgtgaccctgcactgcaccaacctg
aagaacgccaccaacaccaagagcagcaacttggaaaggagatggaccgcggcagatcaagaactgc
agcttcaaggtgaccaccaggcatccgcaacaagatgcagaaggagtacgcctgttctacaagctg
gacgtggtgcccatcgacaacacgacaacaccagctacaagctgtatcaacttgcacaccagcgtgatc
acccaggcctgcccccaaggtgagcttcgagccatccccatccactactgcgcggccggccgttc
gccatcctgaagtgcaacgacaagaagttaacggcagcggccctgcaccaacgtgagcaccgtg
cagtgcacccacggcatcggccctgtggtgagcaccagctgtgtgaaacggcagcctggccag
gagggcgtggtgatccgcagcggagaacttcccgacacaacgccaagaccatcatcgtgcagctgaag
gagagcgtggagatcaactgcacccggcccaacaacaacacccgcacccatcgcc
ggccgcgccttctacgcccacccgcacatcatcgccgcacatccgcaggccactgcaacatcagc
ggcggagaagtggaaacaacacccctgaagcagatcgtgaccaagctgcaggcccgatcggcaacaag
accatcgtgttcaagcagagcagccgcaccccgagatcgtgtgcacagcttcaactgcggc
ggcgagttcttctactgcaacacgcacccagcttcaacagcacctggaacaaccatcgcc
aacaacaccaacggcaccatcaccctgcctgcgcacatcaaggcactcaaccgcggcaggag
gtggcaaggccatgtacgcggccggcaaggagatcagaacaccacccgagatcttccggccggc
ctgctgtgacccggcggcggccggcaaggagatcagaacaccacccgagatcttccggccggc
ggcgacatgcgcgacaactggcgcagcgcagctgtacaaggatcggatcgagccctg
ggcggtggcccccaccaaggccatcagcagcgtggcagagcggagaaggcgcgtgaccctggc
ccatgttccctggcttccctggcgccggcggcggcggcggcggcggcggcggc
gtgcaggcccgccagctgtgagcggcatcgtgcagcagcagaacaacccctgtgcgcgc
gcccagcagcacctgcgcagctgaccgtgtgggcacatcaagcagctgcaggccgcgtgctggc
gtggagcgtcacctgtggatggaccaggcggcggcggcggcggcggcggcggc
accacccggccgtgcctggaaacggccagctggagcaacaagagcctggaccagatctgg
acctggatggatggggcggcggcggcggcggcggcggcggcggcggcggc
agccagaaccaggcaggagaagaacgcggcggcggcggcggcggcggcggc
aactgggtcgacatcagcaagtggctgtggatcatctaactcgag

FIG. 29

(SEQ ID NO:42)

41 / 131

gp140.mut7.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggtctgtgtgtgtgtggaggcagtc
ttcgrrtcgccccagcgccgtggagaagactgtgggtgaccgttactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacagt
tggccaccacccacgcctgcgtgccaccgaccccaaccccaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggaaagaacaacatgtggagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagactgtgacccctgtgcgtgaccctgactgcaccaacctg
aagaacgccaccaacaccaagagcagcaacttgaaggagatggaccgcggcagatcaagaactgc
agcttcaaggtggcgccggcaagctgtatcaactgcacacaccagcgtgatcacccaggcctgcccc
aaggtgagttcgagccatccccatccactactgcgc(cccggccggttcgcacatctgaagtgc
aacgacaagaagtcaacggcagcggccctgcaccaacgtgagcaccgtgcagtgcacccacggc
atccgccccgtgttgcggcaccctgactgtgttgcacggcagcgtggccgaggaggggcgtggatc
cgcagcggagaacttaccgacaacgcaccaagaccatcatcgtgcagctgaaggagagcgtggagatc
aactgcacccggcccaacaacaacacccogcaagagcatcaccatcgccccggccgcgccttctac
gccacccggcgacatcatcgccgacatccggccaggcccactgcacatcagcggcgagaagtggAAC
aacaccctgttgcggcaccatcgatcgttgcacagctgcacgcgttcaactgcggcggcgagtttttctac
cagaggcggcgccgaccccgagatcgtgtatgcacagcttcaactgcggcggcgagtttttctac
tgcaacagcaccctggatcaacagcaccctggatcaacacaccatcgcccccaacaacaccaacggc
accatcaccctggccatcgccgcatcaagcagatcatcaaccctgtggcaggagggtggcaaggccatg
tacgccccccatccggccggccagatccgctgcagcggcaacatcaccggcctgtgtgaccctgc
gacggccggcaaggagatcagcaacaccaccgagatcttccggccccggggggcggcgcacatgcgcac
aactggcgcaagcgagctgtacaagtacaagggtggatggatcgagccctggcgtggccccccatg
aaggccatcagcagcgtggatgcagagcggagaagagcgcggcgtgaccctggcgcacatgttccctggc
ttccctggcgccggccggcagcaccatggcgcccccggcagctgaccctgacccgtgcaggccccggc
ctgtgtggcgcatcgatcgagcggcagaacaaccctgtgcggccatcgaggcccagcagcacctg
ctgcagctgaccctgtggggcatcaagcagctgcaggcccggcgtgtggccgtggagcgcttacctg
aaggaccagcagctgtggatctggggctgcagcggcaagctgtatctgcaccaccgcgtgccc
tggaaacgcggcagctggatgcggcaacaagagcctgggaccagatctggaaacaacatgacccatcg
gaggcgcggatcggatcggatcggatcggatcggatcggatcggatcggatcggatcggatcggatc
gagaagaacgagcaggagctgtggagctggatcggatcggatcggatcggatcggatcggatcggatc
agcaagtggatcggatcggatcggatcggatcggatcggatcggatcggatcggatcggatcggatc

FIG. 30
(SEQ ID NO:43)

42 / 131

gp140.mut7.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtctgtgtggaggcagtc
ttcgtttgcggccagcggcgtggagaagctgtgggtgaccgtgtactacggcgtggccgtgtggaaag
gaggccaccaccaccctgttctgcgcacgcacgccaaggcctacgcacaccgagggtgcacaacgtg
tggggccacccacgcctgcgtgcccacccgaccccaaccccccaggagatgtgtggagaacgtgacc
gagaacttcaacatgtggaaagaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccctgtgtggcgccggcaactgccagacc
agcgtatcaccgcggccatccaaagggtgagcttcgagccatccccatccactactgcgcggcc
gcggcgttcgcgcatectgaagtgcacacgacaagaattcaacggcagcggccctgcaccaacgtg
agcacccgtgcagtgcacccacggcatccgcggccatgtgtggtagcaccaggctgtgtgaacggcagc
ctggccgaggaggggcggtggatccgcagcggagaacttcacccgacaaacgcacactgt
cagctgaaggagagcgtggagatcaactgcacccggcccaacaacaacacccgcaagagcatcacc
atcgccccccggccgcgccttctacgcacccggcgaatcatcgccgacatccgcaggcccactgc
aacatcagcggcagaagtggaaacaacaccctgaagcagatcgtgaccaagctgcaggcccaggttc
ggcaacaagaccatcgttcaagcagagcagcggcgccgaccccgagatcgtatgcacagcttc
aactgcggcggcgagttcttactgcaacagcaccaggctgttcaacagcacctggaaacaacacc
atcgcccccaacaacaccaacggcaccatcaccctgcgcgcacatcaagcagatcatcaaccgc
tggcaggagggtggcaaggccatgtacgcggccatccgcggccagatccgcgcaggcagcaac
atcaccggcctgtgtggccatgcaccccgcaaggagatcagcaacaccaccggagatcttccgc
cccgccggcgccgacatgcgcgacaaactggcgcagcggagctgtacaagtacaagggtgaagatc
gagccctggcggtggccatccaaaggccatcagcagcgtggtagcagagcggagaagagcggccgt
accctggcgccatgttccctgggcttccctggcgccgcccggcagcaccatggcgcccgacgcctg
accctgaccgtgcaggcccgccagctgtgtggagccatcgtgcagcggcagaacaacctgtgcgc
gcacatcgaggcccacgcacccgtgcagcgtgaccgtgtgggcatcaagcagctgcaggcccgc
gtgctggccgtggagcgtacctgaaggaccagcagctgtggcatctggggctgcagcggcaag
ctgatctgcacccaccgcgtgcccctggaaacgcggcagctggagcaacaagagcgtggaccagatctgg
aacaacatgacccatggatggagtgggagcgcgcagatcgcacaactacaccaacctgtatctacaccctg
atcgaggagagcagaaccaggcaggagaagaacgcggcagggagctgtggagcgtggacaagtggcc
agcctgtggaaactggatcgcacatcagcaagtggctgtggatcatctaactcgag

FIG. 31

(SEQ ID NO:44)

gp140.mut8.modsF162

gaattcggcaccatggatgcaatgaagagagggctctgtgtgctgtgtgtggaggcagtc
ttcgccccccagcgccgtggagaagctgtgggtgaccgttactacggcgtgcccgtgtggaaag
gaggccaccaccacccctgttctgcgcagcgcacggccatcgcacaccggaggtgcacaacgtg
tggccaccaccacgcgtgcgtgcccaccgaaaaaccccaaccccaaccccaaggagatgtgtggagaacgtgacc
gagaacttcaacatgtggaaaacaacatggtgaggcagatgcacgaggacatcatcgcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacctg
aagaacgccaccaacaccaagagcagcaacttggaaaggagatggaccgcggcgagatcaagaactgc
agcttcaaggtgaccaccaggcatccgcacaacaagatgcagaaggagtacccctgttctacaagctg
gacgtgggtgcccacgcacaacgcacaacaccaggatcacaagctgtatcaactgcacacaccaggctgatc
acccaggccgcggccatccgcggccatccccatccactacttgcgcggccggccggcttc
gcacatcctgaagtgcacacgacaagaagtcaacggcagcggccctgtgcaccaacgtgaggcacccgtg
cagtgcacccacggcatccgcggccgtggtagccaggatcgtgtgcacccggccatccactacttgcgcggccggcc
gagggcgtggtagccgcggccatccgcggccatccgcggccatccactacttgcgcggccggcc
gagagcgtggagatcaactgcacccggcccaacaacacaacccgcacggccatccactacttgcgcggcc
ggccgcgccttctacgcacccggccatccgcggccatccgcgcacatccgcggccatccgcggcc
ggcgagaagtggaaacaacacccttgcggccatccgcggccatccgcggccatccgcggcc
accatcgtgttcaaggcagccggccatccgcggccatccgcggccatccgcggcc
ggcgagttcttctactgcacacgcacccaggatcgttcaacacgcacccatccgcggcc
aacaacaccaacggccatccgcggccatccgcggccatccgcggccatccgcggcc
gtggggcaaggccatgtacgcggccatccgcggccatccgcggccatccgcggcc
ctgcgtgcgtacccggccatccgcggccatccgcggccatccgcggccatccgcggcc
ggcgacatcgcgcgacacactggccgcacggccatccgcggccatccgcggcc
ggcgccatccgcggccatccgcggccatccgcggccatccgcggccatccgcggcc
ccatccgcggccatccgcggccatccgcggccatccgcggccatccgcggcc
gtgcggccatccgcggccatccgcggccatccgcggccatccgcggccatccgcggcc
gcccaggcgcacccatccgcggccatccgcggccatccgcggccatccgcggcc
gtggggccatccgcggccatccgcggccatccgcggccatccgcggccatccgcggcc
accacccgcgtgccttggaaacgcgcggccatccgcggccatccgcggccatccgcggcc
acccggatggatggggccatccgcggccatccgcggccatccgcggccatccgcggcc
agccagaaccaggcaggagaagaacgcacccatccgcggccatccgcggccatccgcggcc
aactgggttgcacatccgcggccatccgcggccatccgcggccatccgcggccatccgcggcc

FIG. 32

(SEQ ID NO:45)

44 / 131

gp140.mut8.modSF162.delv2

gaattcgccaccatggatcaatgaagagagggctctgtgtgtgtgtgtggagcagtc
ttcgccccagcgcgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
gaggccaccaccacccctgttctgcgccagcgcacgccaaggcctacgacaccgagggtgcacaacgtg
tggccaccacccacgcctgcgtgcccaccgcaccccaaccccccaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggaaaacaacatggtgaggcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacctg
aagaacgcccaccaacaccaagagcagcaacttggaaaggagatggaccggggcggatcaagaactgc
agcttcaaggtggcgccggcaagctgtatcaactgcacaccagcgtatcaccaggcgtggccccc
aaggtgagcttgcagccatccccatccactactgcgccccccgggttcgcctatcctgaagtgc
aacgacaagaagttcaacggcagcggccctgcaccaacgtgagcaccgtgcagtgcacccacggc
atccgccccgttgtgagcaccagctgcgtgaaacggcagccctggccgaggaggcgtgtgtgatc
cgcagcggagaacttcaccgcacaacgccaagaccatcatcgtgcagctgaaggagagcgtggagatc
aactgcacccgccccacaacaacaacacccgcagagacatcaccatcgccccggccgcctctac
gccacccggcgcacatcatcggcgcacatccgcccaggccactgcacacatcagcggcggagaagtggaaac
aacacccctgaagcagatcgtgaccaagctgcagggccagttggcaacaagaccatcgtgttcaag
cagagcagcggcggcgcaccccgagatcgtgatgcacagcttcaactgcggcggcggagttcttctac
tgcaacagcaccagctgtcaacacgcaccttggaaacaacaccatcgcccccaacaacaccaacggc
accatcacccctgcccgtccgcataaaggcagatcatcaaccggstoggcaggaggtggcaaggccatg
tafcgccccccatccgcggccagatccgcgtgcagcagcaacatcaccggcctgcgtgcaccgc
gacggcggcaaggagatcagcaacaccaccgagatcttccgcggccggcggcgcacatgcgcgac
aactggcgcagcggagctgtacaagggtggtaagatcggccctggcgtggccccccacc
atcgccatcagcagcgtgtgcagagcggagaagagcgcgcgtgcaccctggcgcacatgttctggc
ttcctggcgcggccggcagcaccatggcgcggcgcgcgtgcaccctggcgtgcaggcccgccag
ctgcgtgagcggcatcgtgcagcagcagaacaacctgcgcgcgcacatcgaggcccgagcgcaccc
ctgcagctgaccgtgtggcatcaagcagctgcaggcccggtgcgtggcgtggagcgcgtacctg
aaggaccagcagctgtggcatctgggtgcagcggcaacgtgcgtgcaccaccgcgtgccc
tggaaacgcgcagctggagcaacaagagccggaccagatcgtggaaacaacatgacccctgtacac
gagcgcgcagatcgcacaactacaccaacctgtatcgtggaggagagccagaaccacagc
gagaagaacgcgcaggagctgtggagctggacaagtggccagcctgtggaaactgggttgcacatc
agcaagtggctgtgttacatctaactcgag

FIG. 33
(SEQ ID NO:46)

45 / 131

gp140.mut8.modSF162.delV1V2

gaattcgccaccatggatcaatgaagagagggctctgtgtgtgtgtgtggaggcagtc
ttcgttcgcggccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccggtgaaag
gaggccaccaccaccctgttctgcgcacgcacggccatcgacaccggaggtgcacaacgtg
tggccaccacgcctgcgtgcccaccgaccggaccggaccggaggatcgtgtggagaacgtgacc
gagaacctcaacatgtggaaaacaacatggtgaggcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccggccatgtgcgtggccggcaactgccaagacc
agcgtgatcacccaggcctgccccaaagggtgagctcgagccatccccatccactactgcgcccccc
gccggcttcgcacatcctgaagtgcacacaaaggatcacaacggcagggggccctgcaccaacgtg
agcacccgtgcagtgcacccacggcatccggccctgtggtgagcaccaggatcgtgtgaacggcagc
ctggccgaggaggcgtggatccgcagcgagaacttcaccgcacaacgccaagaccatcatcgatc
cagctgaaggagagcgtggagatcaactgcacccggccaaacaacaacacccgcacagagcatcacc
atcgccccggccgcgccttctacgcacccggcagatcatcgccgacatccgcaggccactgc
aacatcagcggcggagaagtggaaaacaacaccctgaagcagatcgtgaccaagctgcaggcccagtgc
ggcaacaagaccatcgtgttcaagcagcagcagcggccgcgaccggcagatcgtgatgcacagcttc
aactgcggcggcggagttttctactgcacacaggcaccctgatcgttcaacagcaccatggaaacaacacc
atcgcccccaacaacaccaacggcaccatcacccctggccgcataaggcagatcatcaaccgc
tggcaggagggtggcaaggccatgtacggccatccgcgaggccatccgcgtgcagcagcaac
atcaccggccatgtgtgcgtgcacccgcgacggccggcaaggagatcagcaacaccaccggagatctccgc
cccgccggccggcagatcgcgcacaactggcgcagcgcagatcgttacaagggtggtaagatc
gagccctggcgtggcccccaccatcgcacatcagcagcgtggcagcagcggagaagcgcggcgtg
accctggccatgttctggcttctggccggccggcagcaccatggccggccgcagccctg
accctgaccctgcaggccgcgcagatcgtgtgcgcgcagcagaacaacccgtgcgc
gcacatcgaggcccagcagcaccctgcgcagatcgtgcaccgtgtggggcataaggcagatcgcaggccgc
gtgcgtggccgtggagcgtacatcgttgcaggaccagcagatcgtgtggcatactggggcgtgcagcggcaag
ctgatctgcaccaccgcgcgtgc
aacaacatgacatggatggatggagatggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
atcgaggagagcagaaccaggcaggagaagaacggcaggagatcgtggagatcgtggacaatggcc
agcctgtggactggatcgacatcagcaactgtggatgtggatcatctaactcgag

FIG. 34
(SEQ ID NO:47)

46 / 131

gp160.modSF162

FIG. 35
(SEQ ID NO:48)

47 / 131

gp160.modsF162.delv2

gaattcgccaccatggatcaatgaagagagggctctgtgtgtgtgtgtggagcagtc
 ttctgtttcgcccagcggcgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtgaaag
 gaggccaccaccaccctgttctgcgccagcacaaggcctacgacaccggaggtgcacaacgtg
 tggccaccacgcctgcgtgcccaccgaccccaaccccccaggagatcgtgtggagaaacgtgacc
 gagaacttcaacatgtgaaagaacaacatggtgagcagatgcacgaggacatcatcgcctgtgg
 gaccagagcctaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacctg
 aagaacgcaccaacaccaagagcagcaacttggaaaggagatggacccggcgagatcaagaactgc
 agcttcaagggtgggcggcaagctgtatcaactgcacacaccagcgtgatcacccaggcctgcccc
 aaggtgagcttgcagccatccccatccactactgcgcggccgcggcttcgcctgaagtgc
 aacgacaagaagttcaacggcagcggccctgcaccaacgtgagcaccgtgcagtgcacccacggc
 atccgcggccgtggttagcaccctgcgtgtgaacggcagcctggccgaggagggcgtggatc
 cgcagcggagaacttcacccgacaacgcacaagaccatcatcgtgcagctgaaaggagagcgtggagatc
 aactgcacccggccccaacaacaacacccgcagatcaccatcgccggccgcgccttctac
 gccacccggcgcacatcatcgccgacatccgcaggcccactgcacacatcagcggcgagaagttggaaac
 aacaccctgtaaagcagatcgtgaccaagctgcaggcccagttcggtacaagaccatcgtgttcaag
 cagagcagccggccgcaccccgagatcgtgtatgcacagcttcaactgcggccggcgtggatcttctac
 tcaacacgcacccagctgtcaacgcacccatggaaacaacaccatcgccggccacaacaccacggc
 accatcaccctgcctgcgcataaggcagatcatcaaccgcgtggcaggaggtggcaaggccatg
 tacggcccccacatccgcggccagatccgcgtgcagcagcaacatcaccggcctgcgtgtgacccgc
 gacggccggcaaggagatcagcaacaccaccaggatcttccgcggccgcggcgcacatgcgcac
 aactggcgagcagctgtacaagtacaagggtggtaagatcggccctggccatggccatgttccctggc
 aaggccaagcgccgcgtggtgacgcgcgagaagcgcgcgtgacccctggccatgttccctggc
 ttccctggccgcggccgcgcacccatggccgcgcacccatggccgcgcacccatggccatgttccctggc
 ctgcgtgagccgcacgcgcacgcacccatgtgcgcgcacatcgaggcccagcagcaccctg
 ctgcagctgaccgtgtggggcatcaaggcagctgcaggcccgcgtgcgcgcacccatggagcgcacccatg
 aaggaccagcagctgtggcatctggggcgtgcagccgcacccatgtggccatgcgcacccatggcc
 tggaaacgcgcagctggagcaacaagagcctggaccagatctggaaacaacatgacccatggatggagttgg
 gagcgcgagatcgcacaactacaccaacccatgtacccctgatcgaggagagccagaaccagcag
 gagaagaacgcgcaggagctgtggagctggacaagtgccgcgcacccatgtggactggatccgcacatc
 agcaagtggctgtggatcatcaagatcttcatcatgatcgtggccgcgcacccatgtggccatgcgcac
 gtgttcccgatcgtgcgcacccatgtgcgcgcacccatgtggccgcgcacccatgtggccatgcgcac
 cgccatccgcgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccatgcgcac
 cgccgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccatgcgcac
 tgcgcgcgcgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccatgcgcac
 ggccgcgcgcgcgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccatgcgcac
 aagaacagcgccgtgacgcgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccgcgcac
 atcgaggtggccgcgcacccatgtggccgcgcacccatgtggccgcgcacccatgtggccgcgcac
 gagcgcgcgcgcgcacccatgtgtaaactcgag

FIG. 36
(SEQ ID NO:49)

48 / 131

gp160.modSF162.delV1V2

FIG. 37

(SEQ ID NO:50)

49 / 131

gp120wtUS4

ACAAACAGTCTTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG
CAACCACCACTCTGTTTGTGCATCAGATGCTAAAGCATAACAAAGCAGAGGC
ACATAACGTCTGGCTACACATGCCGTGTACCCACAGACCCCAACCCACAG
GAAGTAAATTAAACAAATGTGACAGAAAATTAAACATGTGGAAAAATAACA
TGGTGGAACAGATGCATGAGGATATAATCAGTTATGGGATCAAAGCCTAAA
GCCATGTGTAAAATTAAACCCACTCTGTGTACTTTAAATTGTACTGATAAGT
TGACAGGTAGTACTAATGGCACAAATAGTACTAGTGGCACTAATAGTACTAG
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGAAAAGATGCCA
GAAGGAGAAATAAAAAGTCTCTTCAATATCACCACAAAGTGTAAAGAGATA
AAGTGCAGAAAGAATATTCTCTTCTATAAAACTTGTAGTGTACCAATAGAT
AATGATAATGCTAGCTATAGATTGATAAAATTGTAAATACCTCAGTCATTACACA
AGCCTGTCCAAGGTATCTTGAACCAATTCCCACATTATTGTGCCCGG
CTGGTTTGCATTCTAAAGTGTAAAGATAAGAAGTTCAATGGAACAGGACC
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAAGTAGTA
TCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
GATCTGAAAATTTCACAGACAATGCTAAACCATAATAGTACAGCTGAATGA
ATCTGTAGAAATTAAATTGTATAAGACCCAAACAATAACAAAGAAAAAGTATA
CATATAGGACCAGGGAGAGCATTATGCAACAGGTGATAATAGGAGACA
TAAGACAAAGCACATTGTAACATTAGTAAAGCAAATGGACTAACACTTAA
ACAGATAGTTGAAAATTAAAGAGAACAAATTGGGATAATAAAAACAATAATC
TTAATTCACTCAGGAGGGACCCAGAAATTGTATTCAACAGTTAATTG
TGGAGGGAAATTCTATTGTAATACATCACAACATTAAATAGTACCTGGA
ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC
ATGCAGAATAAGACAAATTATAACATGTGGCAAGAAGTAGGAAAAGCAAT
GTATGCCCTCCCATCAGAGGACAAATTAAATGTCATCAAATATTACAGGG
CTGCTATTAACTAGAGATGGTGGTACTAACATAATAGGACGAACGACACCG
AGACCTTCAGACCTGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT
TATATAATATAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCCACCA
GGCAAAGAGAAGAGTGGTGCAAAGAGAGAGAAAAGA

FIG. 38
(SEQ ID NO:51)

50 / 131

gp140wtUS4

ACAACAGTCTTGTGGTCACAGTCTATTATGGGTACCTGTGTGGAAAGAAG
 CAACCAACCCTCTGTTTGTGCATCAGATGCTAAAGCATACAAAGCAGAGGC
 ACATAACGTCTGGCTACACATGCCTGTACCCACAGACCCCACAG
 GAAGTAAATTAAACAAATGTGACAGAAAATTAAACATGTGGAAAAATAACA
 TGGTGGAACAGATGCATGAGGATATAATCAGTTATGGATCAAAGCCTAAA
 GCCATGTGAAAATTAAACCCACTCTGTTACTTTAAATTGTACTGATAAGT
 TGACAGGTAGTACTAATGGCACAAATAGTACTAGTGGCACTAATAGTACTAG
 TGGCACTAATAGTACTAGTACTAATAGTACTGATAAGTGGAAAAGATGCCA
 GAAGGAGAAAATAAAACTGCTTTCAATATCACCACAAGTGTAAAGAGATA
 AAGTGCAGAAAGAATATTCTCTTCTATAAAACTGATGTAGTACCAATAGAT
 AATGATAATGCTAGCTATAGATTGATAAAATTGTAATAACCTCAGTCATTACACA
 AGCCTGTCAAAGGTATCTTGAAACCAATTCCACATCATTATTGTGCCCGG
 CTGGTTTGCATTCTAAAGTAAAGATAAGAAGITCAATGGAACAGGACC
 ATGAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAGTAGTA
 TCAACTCAACTGCTTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
 GATCTGAAAATTTCACAGACAATGCTAAACCATAATAGTACAGCTGAATGA
 ATCTGTAGAAATTAAATTGTATAAGACCCAACAATAATACAAGAAAAAGTATA
 CATATAGGACCAGGGAGAGCATTATGCAACAGGTGATAATAAGGAGACA
 TAAGACAAGCACATTGTAAACATTAGTAAAGCAAACAGGACTAACACTTAA
 ACAGATAGTGAAAAATTAAAGAGAACAAATTGGGATAATAAAACAATAATC
 TTTAATTCACTCTCAGGAGGGACCCAGAAATTGTATTTCACAGTTAATTG
 TGGAGGGGAATTTCATTTGTAATACATCACAATTAAATAGTACCTGGA
 ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC
 ATGCGAAATAAGACAAATTAAACATGTGGCAAGAAGTAGGAAAAGCAAT
 GTATGCCCTCCCATCAGAGGACAATTAAATGTTCATCAAATATTACAGGG
 CTGCTATTAACTAGAGATGGTGGTACTAACATAATAGGACGAACGACACCG
 AGACCTTCAGACCTGGGGAGGAAACATGAAGGACAATTGGAGTAGCACCCACCA
 TATATAAAATATAAAAGTAGTAAGAACATTGAACCAATTAGGAGTAGCACCCACCA
 GGCAGAGAGAACAGTGGTCAAAGAGAGAAAAGAGCAGTGGACTAGGAG
 CTTTGTTCATTGGGTTCTGGGAGCAGCAGGAAGCAGTATGGCGCAGCGTC
 AGTGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCACAG
 CAGAACAAATTGCTGAGAGCTATTGAGGCACAGCATCTGTTGCAACTCA
 CGGTCTGGGCATCAAACAGCTCCAGGCAAGAATCTGGCTGTGGAAAGATA
 CCTAAAGGATCAACAGCTCTAGGGATTGGGTTGCTCTGGAAAACCTCAATT
 GCACCAACTGTGCCATTGGAACTCTAGTTGGAGTAATAAAATCTGACTGAG
 ATTGGGATAATATGACCTGGATGGAGTGGAAAGAGAAATTGGCAATTATA
 CAGGCTTAATATACAATTAAATTGAAATAGCACAAACAGCAAGAAAAGAA
 TGAACAAAGAATTATTGGAATTAGACAAGTGGCAAGTTGTGGATTGGTT
 GATATAACAAACTGGCTGTGGTATATA

FIG. 39
(SEQ ID NO:52)

51 / 131

gp160wtUS4

ACAAACAGTCTTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG
CAACCACCACTCTGTTTGTCATCAGATGCTAAAGCATACAAAGCAGAGGC
ACATAACGTCTGGCTACACATGCCTGTGTACCCACAGACCCCAACCCACAG
GAAGTAAATTAAACAAATGTGACAGAAAATTAAACATGTGGAAAAATAACA
TGGTGGAACAGATGCATGAGGATAATACTAGTTATGGGATCAAAGCCTAAA
GCCATGTGTAAAATTAAACCCACTCTGTGTACTTTAAATTGTACTGATAAGT
TGACAGGTAGTACTAATGGCACAAATAGTACTAGTGGCACTAATAGTACTAG
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA
GAAGGAGAAATAAAAATGCTCTTCAATATCACCACAAGTGTAAAGAGATA
AAGTGCAGAAAGAATATTCTCTTCTATAAAACTTGTATGTAGTACCAATAGAT
AATGATAATGCTAGCTATAGATTGATAAAATTGTAAATACCTCAGTCATTACACA
AGCCTGTCCAAGGTATCTTGAACCAATTCCATACATTATTGTGCCCGG
CTGGTTTGCATTCTAAAGTGTAAAGATAAGAAGTTCAATGGAACAGGACC
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAAGTAGTA
TCAACTCAACTGCTGTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
GATCTGAAAATTTCACAGACAATGCTAAACCATAATAGTACAGCTGAATGA
ATCTGTAGAAATTAAATTGTATAAGACCCACAATAATAACAAGAAAAAGTATA
CATATAGGACCAGGGAGAGCATTATGCAACAGGTGATAATAAGGAGACA
TAAGACAAGCACATTGTAACATTAGTAAAGCAAATGGACTAACACTTGA
ACAGATAGTGAAGAAATTAAAGAGAACATTGGGATAATAAAACAATAATC
TTAATTCACTCCTCAGGAGGGACCCAGAAATTGTATTCACTGTTAATTG
TGGAGGGAAATTCTATTGTAATACATCACAATTAAATAGTACCTGGA
ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC
ATGCAGAATAAGACAAATTATAAACATGTGGCAAGAAGTAGGAAAAGCAAT
GTATGCCCTCCCATCAGAGGACAAATTAAATGTTCATCAAATTACAGGG
CTGCTATTAACTAGAGATGGTGGTACTAACATAATAGGACGAACGACACCG
AGACCTTCAGACCTGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT
TATATAAAATATAAAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCCACCC
GGCAAAGAGAACAGTGGCAGAGAGAAAAGAGCAGTGGACTAGGAG
CTTGTCAATTGGTTCTGGAGCAGCAGGAAGCAGTGGACTAGGAG
AGTGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAACAG
CAGAACAAATTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTCA
CGGTCTGGGCATCAAACAGCTCCAGGCAAGAACCTGGCTGTGGAAAGATA
CCTAAAGGATCAACAGCTCTAGGGATTGGGTTGCTCTGGAAAAGTCAATT
GCACCACTACTGTGCCTTGGAACTCTAGTGGAGTAATAATCTGACTGAG
ATTGGATAATATGACCTGGATGGAGTGGAAAGAGAAATTGGCAATTATA
CAGGCTTAATATACAATTAAATTGAAATTAGACAAGTGGCAGTTGTGGAAATTGGTT
TGAACAAGAACATTGGAAATTAGACAAGTGGCAGTTGTGGAAATTGGTT
GATATAACAAACTGGCTGTTATATAAGAATATTCAATGATAGTAGGAG
GCTTGATAGGTTAAGAATAGTTTGTACTTTCTATAGTGAATAGAGTT
AGGCAGGGACTCACCAATATCATTGCAGACCCGCCTCCAGCTCAGAGGG

FIG. 40A

(SEQ ID NO:53)

52 / 131

GACCCGACAGGCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGA
GACAGATCCAATCGATTAGTCATGGATTATGGCACTCATCTGGGACGATCT
GCGGAGCCTGTGCCTCTTCACTACCACCGCTTGAGAGACTTACTCTTGATTG
TAGCGAGGATTGTGGAACCTCTGGGACGCAGGGGGGGAAAGCCCTCAAGTA
TTGGTGGAAATCTCCTGCAGTATTGGAGTCAGGAGCTAAAGAGTAGTGCTGTT
AGTTTGTAAATGCCACAGCAATAGCAGTAGCTGAAGGGACAGATAGGATTA
TAGAAATAGTACAAAGAATTAGAGCTGTAATTACACATACTAGAAGAAT
AAGACAGGGCTTGGAGAGGGCTTACTATAA

FIG. 40B
(SEQ ID NO:53)

53 / 131

gp120.modUS4

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
GTCTTCGTTGCCAGGCCACCACCGTCTGCTGGGTGACCGTGACTACGGCGTCCCCGTG
TGGAGGAGGCCACCACCAACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCAGGC
CCACAACTGTGGGCCACCCACGCCCTGCGTGCACCCCAACCCCCAGGAGGTGAACC
TGACCAACGTGACCGAGAACCTCAAATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCAAGGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG
ACCCCTGAAC TGACCGACAAGCTGACCGGCAGCACCAACGGCACCAACAGCACCGGCAC
AACAGCACCAAGCGGCACCAACAGCACCCAGCACCAACAGCACCCAGGAGAAGATG
CCCGAGGGCGAGATCAAGAACCTGACGCTTCAACATCACCACCGCGTGCACAGGTCAC
GAAGGAGTACAGCCTGTTCAACAGCTGGACGTGGTGCCTCATCGACAACGACAACGCCAGCT
ACCGCCTGATCAACTGCAACACCAAGCGTGTACCCCAGGCCCTGCCCAAGGTGAGCTTCGAGC
CCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCGCCATCCTGAAAGTGCAGGCCAAGAAGT
TCAACGGCACCGGCCCTGCAAGAACCGTGAGCACCGTGCAGTGCACCCACGGCATCCGCC
GTGGTGAACGCCAGCTGCTGCTGAAACGGCAGCCTGGCGAGGAGGAGATCGTGTGCGCTC
CGAGAACCTCACCGACAACGCCAAGACCATCATCGTGCAGCTGAACCGAGTCCGTGGAGATCA
ACTGCATCCGCCCAACAACAAACACCGCTAAGAGCATCCACATCGGCCCGGCCGCGCTTCT
ACGCCACCGCGACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
TGGACCAACACCTCGAGCAGATCGTGGAGAAGTGCAGCAGTTCGGCAACAAACAGAC
CATCATCTTCAACACGAGCAGCGGCCGACCCGAGATCGTGTCCACAGCTTCAACTGCGG
CGCGAGTTCTTCACTGCAACACCAAGCCAGCTGTTCAACAGCACCTGGAACATCAGCGAGGA
GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCATCCGCCAGATCATCA
ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTCCATCCGCCGGCAGATCAAGTGC
AGCAGCAATATTACCGGCCTGCTGCTGACCCCGACGGCGGCACCAACAAACACCGCACCAA
CGACACCGAGACCTTCCGCCCGGGCGGCAACATGAAGGACAACCTGGCGAGCAGCTGT
ACAAGTACAAGGTGGTGCACATCGAGCCCTGGCGTGGCCCCACCCAGGCCAGGCCAGCG
GTGGTGCAGCGCAGAGCGCTAAGATATCGGATCCTCTAGA

FIG. 41
(SEQ ID NO:54)

54 / 131

gp120.mod.US4.del128-194

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGCTGTGG
AGCAGTCTTCGTTGCCAGGCCACCACCGTGTGGTGACCGTGACTACGGCG
TCCCCGTGTGGAAGGAGGCCACCACCCCTGTTCTGCAGCAGCCAAGGCTTAC
AAGGCCGAGGCCAACCGTGTGGGCCACCCACGCCCTGCGTGCCCCACCGAACCCAAACCC
CCAGGAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGG
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG
AAGCTGACCCCCCTGTGCGTGGGGCAGGGAACTGCGAGACCAGCGTGATCACCCAGGC
CTGCCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCGCCGGCTTCG
CCATCCTGAAGTGCAAGGACAAGAAGTCAACGGCACCGGCCCCCTGCAAGAACGTGAGC
ACCGTGCAGTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGG
CAGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAACCTCACCGACAACGCCAAGA
CCATCATCGTGCAGTGAACGAGTCCGAGATCAACTGCATCCGCCCCAACAAAC
ACCGTAAAGAGCATCCACATCGGCCCCGGCGCCCTTACGCCACCGGCGACATCAT
CGGCGACATCCGCCAGGCCCCTGCAACATCAGCAAGGCCAACTGGACCAACACCCCTG
AGCAGATCGTGGAGAAGCTGCGCAGCAGTTGGCAACAAAGACCATCATCTCAAC
AGCAGCAGCGGCGGGGACCCCGAGATCGTGTCCACAGCTCAACTGCGGGCGAGTT
CTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGA
ACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGACATCCGCCAGATCATCAAC
ATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGCCGACGGCGACCAACAACCGCA
CAGCAGCAATATTACCGGCCCTGCTGACCCCGCACGGCGACCAACAACCGCA
CCAACGACACCGAGACCTTCCGCCCCGGCGGCAACATGAAGGACAACCTGGCGCAGC
GAGCTGTACAAGTACAAGGTGGTGCATCGAGCCCCCTGGCGTGGCCCCCACCCAGGC
CAAGCGCCCGTGGTGCAGCGCAGAACGCGTAAGATATCGGATCCTCTAGA

FIG. 42

(SEQ ID NO:55)

55 / 131

gp140.modUS4

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGTGGAGCA
GTCTCGTTGCCAACGCCACCGTGTGCTGGGTGACCGTGTACTACGGCGTCCCCTG
TGGAAGGAGGCCACCACCCACCGTGTGCTGCCAGCGACGCCAAGGCTAACAGGCCAGGC
CCACAACGTGTGGGCCACCCACGCCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC
TGACCAACGTGACCGAGAACTTCAAATGTGGAAGAACAAACATGGTGGAGCAGATGCAATGAG
GACATCATCAGCCTGTGGGACCAAGAGCCTGAAGCCTGCGTGAAGCTGACCCCCCTGCGTG
ACCCCTGAACIGACCGACAAGCTGACCGCAGCACCAACGGCACCAACAGCACCAAGCGGAC
CACACGACCCAGCGGCACCAACAGCACCAACAGCACCGACAGCTGGGAGAAAGATG
CCCGAGGGCGAGATCAAGAACCTGCACTTCAAATCACCACCAAGCGTGCAGCAACAGGTC
GAAGGAGTAACAGCCTGTCTACAAGCTGGACGTGGTGCCTACGACAACGACAACGCCAGCT
ACCGCCTGATCAACTGCAAACACCAGCGTATCACCAAGGCTGCCAACGGTAGCTTCGAGC
CCATCCCCATCCACTACTGCGCCCCCGCCCGCTTCGCCATTCCTGAAGTGCAGGACAAGAAGT
TCAACGGCAACGGCCCTGCAAGAACAGTGAACCGTGCAGTGCACCCACGGCATCCGCC
GTGGTGAACGACCCAGCTGCTGTAACGGCAGCCTGGCCAGGGAGGAATCGTGCCTGC
CGAGAACATTACCGACAACGCCAACACCGCATCTGTGCAAGCTGAACGAGTCCGTGAGATCA
ACTGCATCCGCCCCAACAAACAAACACCGCTAACAGGACATCCACATCGGCCCCGGCC
ACGCCACCGCGACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
TGGACCAACACCCCTGAGCAGATCGTGGAGAACGCTGCGCAGCAGTTCGGCAACAAAGAC
CATCATCTAACAGCAGCAGCGGGGGGACCCGAGATCGTGTCCACAGCTAACATGCC
CGCGAGTTCTACTGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCACCGAGGA
GGTGAACAGAACCAAGGAGAACGACACCATCATCTGCCCTGGCGATCCGCCAGATCAAGTGC
ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTCCATCCGCCAGATCAAGTGC
AGCAGCAATATTACCGGCTGCTGACCCGCGACGGGGACCAACAAACACCGCACCA
CGACACCGAGACCTCCGCCCGGGCAACATGAAGGACAATGGCGCAGCGAGCTGT
ACAAGTACAAGGTGGTGCACATGAGCCCCCTGGCGTGGCCCCACCAAGGCCAACGCC
GTGGTGCAGCGCAGAACGCGCCGTGGCGCCCTGTTATCGGCTTCTGGCG
GCCGGGAGCACCATGGCGCCGCTCCGTACCCGACCGTGCAGGCCAGGCCAGCTGCTGAG
CGGCATCGTCAGCAGCAGAACACCTGCTGCGGCCATCGAGGCCAGCAGCACCTGCTGC
AGCTGACCGTGTGGGATCAAGCAGCTGCAGGCCGATCTGCCCTGGAGCGCTACCTG
AAGGACCAAGCAGCTGCTGGGATCTGGGCTGACGGCAAGCTGATCGACCCACCGT
GCCCTGGAAACAGCAGCTGGAGCAACAAAGAGCCTGACCGAGATCTGGACAACATGACCTG
TGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCATCTACAACCTGATCGAGATCG
CAGAACCAAGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTG
GGAACCTGGTCAACATCACCAACTGGCTGTGGTACATCTAACGATATCGGATCCTCTAGA

FIG. 43

(SEQ ID NO:56)

gp140.mut.modUS4

GAATTGCCACCATGGATGCAATGAAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
GTCTTCGTTGCCAGGCCACCACCGTCGCTGGGTGACCGTGTACTACGGCGTCCCCGTG
TGGAGGAGGCCACCACCAACCTGTTCTGCAGCAGGCCAACGACCCCAACCCCCAGGAGGTGAACC
CCACAACGTGTGGCCACCCACGCCCTGCGTGCCTGCCCCACCGACCCCAACCCCCAGGAGGTGAACC
TGACCAACGTGACCGAGAACATCAACATGTGGAGAACAACTGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACAGAGCCTGAAGGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG
ACCCCTGAACTGCACCGACAAGCTGACCGCAGCACCAACCGCACCAAACGACACCAGCGCAC
CAACAGCACCAGCGCACCAACAGCACCGCACCAACAGCACCGACAGCTGGAGAACAGATG
CCCAGGGCGAGATCAAGAACCTGCACTTCAACATCACCAACAGCGTGCAGCACAGGTGCA
GAAGGAGTAACGGCTGTTCAACAGCTGGACGGTGGCCCATCGACAAACGACAAACGCCAGCT
ACCGCCTGATCAACTGCAACACCGAGCGTGTACCCAGGCCCTGCCCCAAGGTGAGCTTCGAGC
CCATCCCCATCCAACACTGCCCCCGGCTCGCCATCCTGAAGTGAAGGACAAGAAGT
TCAACGGCACCGGCCCTGCAAGAACCGTGGAGCACCGTGCAGTGCACCCACGGCATCGGCCCC
GTGGTAGCACCCAGCTGCTGTAACGGCAGCTGGCCAGGGAGAGATCGTGTGCGCTC
CGAGAACTTACCGACAACGCCAACGACCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCA
ACTGCA TCGCCCCAACAAACACCGCTAACAGCATCCACATCGGCCCCGGCGCCCTCT
ACGCCACCGCGACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
TGGACCAACACCCCTGAGCAGATCGTGGAGAACGCTGCCAGCGAGCTTCCGCAACAAAGAC
CATCATCTTCAACAGCAGCGCGCCGAGACCCCGAGATCGTGTCCACAGCTCACTCGG
CGGCAGTTCTACTGCAACACCGCCAGCTTCAACAGCACCTGGAACATCACCGAGGA
GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCA
ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTCCGATCCGCCAGATCAAGTGC
AGCAGCAATTACCGGCCCTGCTGACCCCGCAACGGCGGACCCAAACAAACACCGCACCAA
CGACACCGAGACCTTCCGCCCGCGCGCAACATGAAGGACAACGGCGAGCGAGCTGT
ACAAGTACAAGGTGGTGCATCGAGCCCTGGCGTGGCCCAACCCAGGCCAACGCC
GTGGTGCAGCGAGAACAGCGCCGTGGCCTGGCGCCCTGTTCATCGCTCTGGCGCC
GCCGGAGCACCATTGGCGCCGCTCGTGAACCTGACCGTGCAGGCCGCCAGCTGTGAG
CGGCATCGTGCAGCAGCACAACTGCTGCCCATCGAGGCCAGCAGCACCTGCTG
AGCTGACCGTGTGGGCATCAAGCAGCTGCAAGGCCGATCCTGGCGTGGAGCGTACCTG
AAGGACCAAGCAGCTGCTGGCATCTGGGCTGAGCGGAAGCTGATCTGCACCAACCCGT
GCCCTGGAACAGCAGCTGGAGCAACAGAGCCTGACCGAGATCTGGACAAACATGACCTGGA
TGGAGTGGAGCGCGAGATCGGCAACTACACCGGCCCTGATCTACAACCTGATCGAGATCGCC
CAGAACCAAGCAGGAGAACAGGAGAGCTGCTGGAGCTGGACAAAGTGGGCCAGCCTGT
GGAACCTGGTTCGACATCACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 44
(SEQ ID NO:57)

57 / 131

gp140.TM.modUS4

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
GTCTTCGTTGCCAGCGCCACCACCGTCTGCTGGGTGACCGTGACTACGGCGTGCCTG
TGAAGGAGGCCACCAACCACCCCTGTTCTGCGCCAGCGACGCCAAGGCTAACAGGCCAGGC
CCACAACGTGTGGGCCACCCACGCCCTGCGTGCCTGCCACCGACCCAAACCCCCAGGAGGTGAACC
TGACCAACCGTACCGAGAACTTCAACATGTGGAAAGAACAAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCAAGGCTGAAGCCCTGCGTGAAGCTGACCCCTGTGCGTG
ACCTGAACCTGACCGACAAGCTGACCGCAGCACCAACGGCACCAACAGCACCCAGCGGCAC
AACAGCACCAAGCGCACCAACAGCACCAACAGCACCGACAGCTGGAGAAAGATG
CCCGAGGGCGAGATCAAGAACCTGCGACTTCAACATCACCAACCGCGTGCCTGCGACAAGGTGCA
GAAGGAGTAAGCCTGTTCTACAAGCTGGACGTGGTGCCTATCGACAAACGACAACGCCAGCT
ACCGCCTGATCAACTGCAACACCAGCGTATCACCCAGGCCCTGCCAACGGTGAAGCTTCGAGC
CCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCGCCTACCTGAAGTGAAGGACAAGAAC
TCAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGAGTCACCGCAAGGCACTCGCC
GTGGTGAGCACCCAGCTGCTGTAACGGCAGCCTGGCCAGGGAGATCGTGTGCGCTC
CGAGAACCTCACCGACAACGCCAACGACCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCA
ACTGCATCCGCCCCAACAACACAGCGTAAGAGACATCCACATCGGCCCGGCCCTCT
ACGCCACCGCGACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
TGGACCAACACCCCTGAGCAGATCGTGGAGAACGCTGCGAGCAGTTCGGCAACAAACAGAC
CATCATCTCAACAGCAGCGCGGCCAGCCAGATCGTGTCCACAGCTTCAACTGCGG
CGGCGAGTTCTACTGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCACCGAGGA
GGTGAACAAAGCCAAGGAGAACGACACCATCATCTGCCCTGCCGAATCGGCCAGATCATCA
ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGCCATCGCGGCCAGATCAAGTGC
AGCAGCAATATTACCGCCTGCTGTAACCGCGACGGCGGCCAACAAACACCGCACCAA
CGACACCGAGACCTCCGCCCGCGGCCAACATGAAGGACAACCTGGCGCAGCGAGCTGT
ACAAGTACAAGGTGGTGCATCGAGCCCTGGCGTGGCCCTGCCAGGCAAGGCC
GTGGTGAGCGCGAGAACGCGCCGTGGCGCCCTGTTATCGGCTTCTGGCGCC
GCCGGGAGCACCATGGCGCCGCTCCGTGACCGTGCAGGCCAGCTGAG
CGGCATCGTGCAGCAGCAGAACACCTGCTGCGCCATCGAGGCCAGCAGCACCTGCTGC
AGCTGACCGTGTGGGATCAAGCAGCTGCAAGGCCGATCCTGCCGTGGAGCGCTACCTG
AAGGACCAAGCAGCTGCTGGGATCTGGGCTGCAAGCGAACGCTGATCTGCACCAACCGT
GCCCTGGAACAGCAGCTGGAGCAACAGAGCTGACCGAGATCTGGACAAACATGACCTGGA
TGGAGTGGAGCGCGAGATCGGCAACTACACCGCCTGATCTACAACCTGATCGAGATCGC
CAGAACCAAGCAGGAGAAGAACGAGCAGGGAGCTGCTGGAGCTGGACAAGTGGCCAGCCTG
GGAACCTGGTTGACATCAACACTGGCTGTGGTACATCGCACTTCATCATGATCGTGGCG
GCCTGATCGGCCCTGCGCATCGTGTGCGTGCAGCATCGTGTAAAGATATCGGATCCTCTA
GA

FIG. 45
(SEQ ID NO:58)

58 / 131

Gp140modUS4.DV1V2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGC
TGTGTGGAGCAGTCITCGTTGCCAGCGCCACCACCGTGTGGTGGTGACC
GTGTACTACGGCGTGCCCGTGTGGAGGAGGCCACCACCCCTGTTCTGCG
CCAGCGACGCCAAGGCTTACAAGGCCAGGCCAACACGTGTGGGCCACCCA
CGCCTGCGTGCCACCGACCCCCAACCCCCAGGAGGTGAACCTGACCAACGTG
ACCGAGAACTTCAACATGTGGAGAACACAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCAAGAGCCTGAAGGCCCTGCGTGGGCGCCGGCC
AGGCCTGCCCAAGGTGAGCTCGAGGCCATCCCCATCCACTACTGCGCCCC
CGCCGGCTCGCCATCCTGAAGTGAAGGACAAGAAAGTTCAACGGCACCGGC
CCCTGCAAGAACGTGAGCACCGTCAGTGCACCCACGGCATCCGCCCGTGG
TGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCAGGGAGGAGATCGTGC
GCGCTCGAGAACTTCAACGACAAGGCCAACATCATCGTCAGCTGAAC
GAGTCCGTGGAGATCAACTGCATCCGCCAACAACACAACACCGCTAACAGCA
TCCACATCGGCCCGGCCGCTTCTACGCCACCGCGACATCATCGCGA
CATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCTC
GAGCAGATCGTGGAGAACAGCTGCGCAGCAGTTCGGAACAAACAAGACCATC
ATCTTCAACAGCAGCAGCGCGGAGATCGTGTCCACAGCTCA
ACTGCGCGGAGTTCTTCTACTGCAACACCAAGCCAGCTGTTCAACAGCAC
CTGGAACATCACCAGGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT
CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTGGCAAG
GCCATGTACGCCCTCCATCCGCCAGATCAAGTGCAGCAGCAATATTA
CCGGCCTGCTGCTGACCCCGCAGGGCGAACATGAAGGACAACCTGGCAGC
CACCGAGACCTCCGCCGGCGGGCAACATGAAGGACAACCTGGCAGC
GAGCTGTACAAGTACAAGGTGGTGCAGCGCATCGAGGCCCTGGCGTGGCCCG
CCCAGGCCAAGCGCCGCTGGTGCAGCGCAGAACGCGCCGTGGCGTGGCG
GCGCCCTGTTCATCGGCTTCTGGCGCCGGGAGCACCATGGCGCCGC
CTCCGTACCCCTGACCGTGCAGGCCAGCTGCTGAGCGGCATCGCAG
CAGCAGAACACCTGCTGCGCCATCGAGGCCAGCAGCACCTGCTGCAG
TGACCGTGTGGGCATCAAGCAGCTGCAGGCCAGCTGCTGAGCGGCATCG
CTACCTGAAGGACCAAGCAGCTGCTGGCATCTGGGCTGCAAGCGGAAAGCTG
ATCTGCACCAACCGTGCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGA
CCGAGATCTGGGACAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCA
ACTACACCGGCCTGATCTACAACCTGATCGAGATGCCAGAACAGCAGGA
GAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAA
CTGGTTCGACATACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTCTA
GA

FIG. 46

(SEQ ID NO:59)

Gp140modUS4.DV2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTGCTGTGCTGCTGC
 TGTGTGGAGCAGTCTCGTTGCCAGCGCCACCACCGTGTGGGGTGC
 GTGTACTACGGCGTCCCCGTGTGGAAGGAGGCCACCACCCCTGTTGCG
 CCAGCGACGCCAACGGCTACAAGGCCGAGGCCACAAACGTGTGGGCCACCCA
 CGCCTGCGTGCCCCACCGACCCCAACCCCCAGGAGGTGAACCTGACCAACGTG
 ACCGAGAACCTAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCTGCGTAAGCTGACCC
 CCCTGTGCGTGACCCCTGAACCTGACCGACAAGCTGACCGGCAGCACCAACGG
 CACCAACAGCACCGAGCGGCACCAACAGCACCAAGCGGCCACCAACAGCACCAAG
 CACCAACAGCACCGACAGCTGGGAGAAGATGCCGAGGGCGAGATCAAGAA
 CTGCAGCTTCAACATCGCGCCGCCCTGATCAACTGCAACACCAAGCGTG
 ATCACCCCAGGCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACT
 GCGCCCCCGCCCGCTTCGCATCCTGAAGTGCAAGGACAAGAAGTTCAACGG
 CACCGGCCCCCTGCAAGAACGTGAGCACCCTGCAAGGGCAGCCCTGGCGAGGAGGAGA
 CCCGTGGTGAGCACCCAGCTGCTGTGAACGGCAGCCCTGGCGAGGAGGAGA
 TCGTGCTGCGCTCCGAGAACCTAACCGACAACGCCAACATCATCGTGCA
 GCTGAACGAGTCCGTGGAGATCAACTGCACTCCGCCAACAAACAACACGCGT
 AAGAGCATCCACATCGGCCCCGGCCGCCTTCTACGCCACCGGCACATCA
 TCGGCGACATCCGCCAGGCCCAGTCAACATCAGCAAGGCCAACGGACCAA
 CACCCCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAA
 GACCATCATCTCAACAGCAGCAGCGGGCGGCCAGCCGAGATCGTGTCCAC
 AGCTTCAACTGCGGGCGAGTTCTACTGCAACACCAAGGCCAGCTGTTCAA
 CAGCACCTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACAC
 CATCATCCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAGGAGGTG
 GGCAAGGCCATGTACGCCCTGCCAGATCAAGTGCAGCAGCA
 ATATTACCGGCCCTGCTGTGACCCCGCAGGGCGGCCACCAACAAACACCGCAC
 CAACGACACCGAGACCTCCGCCCGCGCGGCCACATGAAGGACAACATG
 GCGCAGCGAGCTGTACAAGTACAAGGTGGTGCAGCAGGCCCTGGCGTG
 GCCCCCACCCAGGCCAGCGCCCGTGGTGCAGCGCAGAACAGCGCCGTG
 GGCCTGGCGCCCTGTTCATCGGCTTCTGGCGCCGGAGCACCATGG
 GCGCCGCCTCCGTGACCTGACCGTGCAGGCCAGCTGCTGAGCGGCAT
 CGTGCAGCAGCAAAACCTGCTGCGGCCATCGAGGCCAGCAGCACCTG
 CTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCAGATCCTGGCG
 TGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGAGCGG
 CAAGCTGATCTGCACCACCAACCGTGCCTGGAACAGCAGCTGGAGCAACAAG
 AGCCTGACCGAGAGATCTGGGACAACATGACCTGGATGGAGTGGAGCGCGAG
 ATCGGCAACTACACCGGCCTGATCTACAAACCTGATCGAGATGCCAGAAC
 AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCC
 TGTGGAACTGGITCGACATCACCAACTGGCTGTGGTACATCTAAGATATCGG
 ATCCTCTAGA

FIG. 47
(SEQ ID NO:60)

60 / 131

Gp140modmutUS4.DV1V2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTTGCTGTGCTGCTGC
TGTGTGGAGCAGTCCTCGTTGCCAGGCCACCACCGTGTGGTGGTGGACC
GTGTACTACGGCGTCCCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCG
CCAGCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTGGGCCACCC
ACGCCTGCGTGCCACCGACCCCCAACCCCCAGGAGGTGAACCTGACCAACGT
GACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGATGA
GGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGGGCCGCC
CAGGCCCTGCCCAAGGTGAGCTTCGAGCCCATAACCAACTACTGCGCCC
CCGCCGGCTGCCATCCTGAAGTGCAAGGACAAGAAAGTCAACGGCACCGG
CCCCCTGCAAGAACGTGAGCACCCTGCAGTGACCCACGGCATCCGCCCGTG
GTGAGCACCAGCTGCTGTAACGGCAGCCTGGCCGAGGAGGAGATCGTGC
TGCCTCCGAGAACTTCAACGACAACGCCAAGACCATCATCGTGCAGCTGAA
CGAGTCCGTGGAGATCAACTGCATCCGCCAACAAACAACACCGCTAAGAGC
ATCCACATCGGCCCCGGCCGCGCCTCTACGCCACCCGACATCATCGGCG
ACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCC
CGAGCAGATCGTGGAGAAGCTGCGCAGCAGTTGGCAACAACAAGACCAT
CATCTTCAACAGCAGCAGCGGGCGACCCCGAGATCGTGTCCACAGCTTC
AACTGCGGCGGCGAGTTCTACTGCAACACCAGCCAGCTGTTCAACAGCA
CCTGGAACATCACCAGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCA
TCCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTGGCAA
GGCCATGTACGCCCTCCATCCGCCAGATCAAGTGCAGCAGCAATATT
ACCGGCTGCTGCTGACCCCGACGGCGACCAACAACACCGCACCAACG
ACACCGAGACCTTCCGCCGGCGGCAACATGAAGGACAACACTGGCGCA
GCGAGCTGTACAAGTACAAGGTGGTGCACGAGCCGACCAACAACACCG
CACCCAGGCCAAGCGCCGCGTGGTGCAGCGCAGAAGAGCGCCGTGGCCT
GGGCGCCCTGTTCATCGGCTTCCCTGGGCCGCCGGAGCACCATGGCGCC
GCCTCCGTGACCTGACCGTGCAGGCCAGCTGCTGAGCGGCGATCGTGC
AGCAGCAGAACACCTGCTGCGGCCATCGAGGCCAGCAGCACCTGCTGCA
GCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCAGCTGGCATCTGGGCTG
CGCTACCTGAAGGACCAAGCAGCTGCTGGCATCTGGGCTGAGCGGCAAGC
TGATCTGACCAACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGC
GACCGAGATCTGGGACAACATGACCTGGATGGAGTGGAGCGCGAGATCG
CAACTACACCGCCCTGATCTACAACACTGATCGAGATGCCAGAACCGAGC
GAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTG
AACTGGTTCGACATACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTC
TAGA

FIG. 48

(SEQ ID NO:61)

61 / 131

gp140.mod.US4.del128-194

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGCTGTGTGG
 AGCAGTCTCGTTGCCAGGCCACCACCGTGTGGTGACCGTGTACTACGGCG
 TGCCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCAGCGACGCCAAGGCTTAC
 AAGGCCGAGGCCACAACTGTGTGGCACCCACGCCTGCGTGCCTGCCCCACCGACCCCAACCC
 CCAGGAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGG
 TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACAGAGCCTGAAGCCCTGCGTG
 AAGCTGACCCCCCTGTGCGTGGGGCAGGAACTGCGAGACCAAGCGTGTACCCAGGC
 CTGCCCAAGGTGAGCTCGAGCCATCCCCATCCACTACTGCGCCCCGCCGGCTTCG
 CCATCCTGAAGTGAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGC
 ACCGTGCAGTGCACCCACGGCATCCGCCGTGGTAGACCCAGCTGCTGTAACCG
 CAGCCTGGCGAGGAGGAGATCGTGTGCGCTCCGAGAACTTACCCACAAGCCAAGA
 CCATCATCGTGCAGTGAACGAGTCCGTGGAGATCAACTGCACTCCGCCAACAAAC
 ACGCGTAAGAGCATCCACATCGGCCCCGGCCGCGCTTCTACGCCACCGCGACATCAT
 CGGCCACATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCTCG
 AGCAGATCGTGGAGAAGCTGCGCAGGAGTTGCAACAACAAGACCATCATCTTCAAC
 AGCAGCAGCGCGGGGAGCCCGAGATCGTGTTCACAGCTTCAACTGCGGCGGAGTT
 CTTCTACTGCAACACCAAGGCCAGCTGTTCAACAGCACCTGGAACATCACCAGGAGGTGA
 ACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCCATCCGCCAGATCATCAAC
 ATGTGGCAAGGAGGTGGCAAGGCCATGTAACGCCCGGCGACGGCCAGATCAAGTG
 CAGCAGCAATATTACCGCCCTGCTGCTGACCCCGGACGGCCAGCAACAACAACCGCA
 CCAACGACACCGAGACCTTCCGCCCCGGCGGCAACATGAAGGACAACACTGGCGCAGC
 GAGCTGTACAAGTACAAGGTGGTGCAGCGAGAAGCGCGCCGTGGGCGTGGCCCCCAGGC
 CAAGCGCCCGTGGTGCAGCGAGAAGCGCGCCGTGGGCGCCCTGTTCATCG
 GCTTCTGGCGCCGCCGGAGCACCATGGCGCCGCGCTCCGTGACCCCTGACCGTGCAG
 GCCCGCCAGCTGCTGAGCGGACATGTCAGCTGACCGTGTGGGCAAGCAGCTGCAGGCCAGC
 GGCCCGAGCACCTGCTGCAAGCTGACCGTGTGGGCAAGCAGCTGCAGGCCAGC
 TCCTGGCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCACTGGGCTGAGC
 GGCAAGCTGATCTGCAACACCACCGTCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT
 GACCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACA
 CGGGCCTGATCTACAACCTGATCGAGATCGCCAGAACCCAGCAGGAGAAGAACGAGCAG
 GAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGGAACCTGGTCAACATCACCACCTG
 GCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 49
(SEQ ID NO:62)

62/131

gp140.mut.mod.US4.del128-194

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGTGTGG
AGCAGTCTTCGTTCGCCCAGGCCACCACCGTGTGGTGAACGTGTACTACGGCG
TGCCCGTGTGGAAGGAGGCCACCACCCCTGTTCTGCGCCAGCGACGCCAAGGCTTAC
AAGGCCGAGGCCACAACGTGTGGCCACCCACGCCCTGCGTGCCTACCGACCCAAACCC
CCAGGAGGTGAACTTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGG
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGGCCTGAAAGCCCTGCGTG
AAGCTGACCCCCCTGTGCGTGGGGCAGGGAACTGCGAGACCAGCGTGTACCCAGGC
CTGCCCAAGGTGAGCTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCG
CCATCCTGAAGTGCAGGACAAGAAGTTCAACGGCACCGGCCCCCTGCAAGAACGTGAGC
ACCGTGCAGTGCACCCACGGCATCCGCCCGTGGTGAAGCACCCAGCTGCTGCTGAAACGG
CAGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAACTTCAACCGACAACGCCAAGA
CCATCATCGTGCAGCTGAAACGAGTCCGTGGAGATCAACTGCACTCCGCCCAACAACAAAC
ACCGTAAGAGCATCCACATCGGCCCCGGCGCCCTTCAACGCCACCGGCAACATCAT
CGCGACATCCGCCAGGCCACTGCAACATCAGCAAGCCAACGGACCAACACCTCG
AGCAGATCGTGGAGAAGCTGCGCAGCAGTTCGGCAACAACAAGACCATCATCTTCAAC
AGCAGCAGCGGGCGACCCGAGATCGTGTTCACAGCTTCAACTGCGCGGCGAGTT
CTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGAAACATCACCGAGGAGGTGA
ACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCAAC
ATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGCCGAGATCAAGTGA
CAGCAGCAATATTACCGGCTGCTGTCACCGCGACGGCGGCAACATGAAGGACAACGGCA
CCAACGACACCGAGACCTCCGCCCGCGGCAACATGAAGGACAACGGCAGC
GAGCTGTACAAGTACAAGGTGGTGCAGCGAGAACAGCGCCGTGGCCTGGCGCCCTGTTCA
CAAGCGCCGCGTGGTGCAGCGAGAACAGCGCCGTGGCCTGGCGCCCTGTTCA
GCTTCCCTGGCGCCGCCGGGAGCACCATGGCGCCCTCCGTGACCTGACCGTGCAG
GCCCGCCAGCTGCTGAGCGGATCGTGCAGCAGCAGAACACCTGCTGCGGCCATCGA
GGCCCGAGCAGCACCTGCTGCACTGCTGAGCGCAGCTGCTGGCCTGGCGCCCTGAGC
TCCTGGCCGTGGAGCGTACCTGAAGGACCAAGCAGCTGCTGGCCTGGCGCCCTGAGC
GGCAAGCTGATCTGACCAACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT
GACCGAGATCTGGGACAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACA
CCGGCCTGATCTACAACCTGATCGAGATGCCAGAACCGCAGGAGAACGAGCAG
GAGCTGCTGGAGCTGGACAAGTGGCCAGCCTGGAACGGTTGACATCACCACCTG
GCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 50
(SEQ ID NO:63)

63/131

gp160.modUS4

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGAGCA
 GTCTTCGTTTGCCTGCCAGGCCACCACCGTGTGGTGACCGTGTACTACGGCGTGTCCCCTG
 TGGAAGGAGGCCACCAACCACCTGTCTGCCAGCAGCCAAGGCTACAAGGCCAGGC
 CCACAACTGTGGGCCACCCACGCCCTGCCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC
 TGACCAACGTGACCGAGAACCTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCAATGAG
 GACATCATCAGCCTGTGGGACCAAGGCCTGAAGCCTGCCGTGAAGCTGACCCCCCTGTGCGTG
 ACCCTGAACCTGCACCGACAAGCTGACCGCAGCACCAACGGCACCAACAGCACCGGGCAC
 CAACAGCACCGCAGCACCAACAGCACCAACAGCACCGACAGCTGGGAGAAGATG
 CCCGAGGGCAGATCAAGAACCTGCAAGCTTCAACATCACCACCGCTGCCGACAAGGTGCA
 GAAGGGATACAGCCTGTTCTACAAGCTGGACGTGGTCCCCATCGACAAACGACAACGCCAGCT
 ACCGCCCTGATCAACTGCAACACCAGCGTGTACCCAGGCCCTGCCCAAGGTGAGCTTCGAGC
 CCATCCCCATCCACTACTGCCCTCCGGCTTCGCCATCCTGAAGTGCAGGACAAGAAGT
 TCAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGCC
 GTGGTGAGCACCCAGCTGCTGTGACGGCAGCCTGGCGAGGAGGAGATCGTGTGCGCTC
 CGAGAACCTACCGACAACGCCAACGACCATCATCGTGCAGCTGAAGGAGTCCGTGGAGATCA
 ACTGCATCCGCCCCAACAAACACGCCAACGCGTAAGAGACATCCACATCGGCCCCGGCGCCTTCT
 ACGCCACCGCGACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
 TGGACCAAACACCTCGAGCAGATCGTGGAGAACGCTGCGCAGCAGTCCGCAACAACAAGAC
 CATCATCTTCAACAGCAGCAGCGCGGCCAACCCAGATCGTGTCCACAGCTTCAACTGCGG
 CGCGAGATTCTTACTGCAACACCAGCCAAGCTTCAACAGCACCTGAAACATCACCGAGGA
 GGTGAACAAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCA
 ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGGCGACGGGGCACCAACAACACCGCACCA
 AGCAGCAATAATTACCGGCTGCTGTGACCCGCAGGGGGCACCAACAACACCGCACCA
 CGACACCGAGACCTCCGCCCGCGGCCAACATGAAGGACAACGGCGCAGCGAGCTGT
 ACAAGTACAAGGTGGTGCACATCGAGCCCCTGGCGTGGCCCCAACCAAGGCCAACGCC
 GTGGTGCGCGAGAACGCGCCGTGGCGCCCTGTCATCGGCTTCTGGCGCC
 GCCGGGAGCACCATGGCGCCGCTCCGTGACCTGACCGTGCAGGCCGCCAGCTGCTGAG
 CGGCATCGTGCAGCAGCAGAACAAACCTGTCGCGCCATCGAGGCCAGCAGCACCTGTC
 AGCTGACCGTGTGGGATCTGGGCTGCAAGCGGCCAGCTGACCGAGATCTGGGACAACATGACCTGGA
 AAGGACGAGCAGCTGGAGCAACAAGAGCTGACCGAGATCTGGGACAACATGACCTGGA
 GCCCTGGAACAGCAGCTGGAGCAACAAGAGCTGACCGAGATCTGGGACAACATGACCTGGA
 TGGAGTGGGAGCGCGAGATCGGCAACTACACCGCCCTGATCTACAAACCTGATCGAGATGCC
 CAGAACCGAGGAGAACGAGCAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCTGT
 GGAACCTGGTTCGACATACCAACTGGCTGTGTCAGCGCAAGCTGATCTGACCC
 GCCTGATCGGCCCTGCCGATCGTGTGCCGTGTCAGCATCGTGAACCGCGGCCAGGGCT
 ACAGCCCCATCAGCGTCAAGACCGCCCTGCCGCCAGCGCGGCCCGACCGCCCCGAGGGC
 ATCGAGGAGGAGGGCGCGAGCGCGACCGCGAGCGCAGCAACCGCCTGGTGCACGGCTGCT
 GGCCCTGATCTGGGACGACCTGCGCAGCCTGTCGCTGCTGTCAGCTACCAACGCCCTGCCGACCT
 GCTGCTGATCGTGGCCGATCGTGGAGCTGCTGGCGCCGCGCTGGAGGGCTGAAGT
 ACTGGTGGAACCTGCTGCAGTACTGGAGGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTC
 AACGCCACCGCAATCGCCGTGGCGAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCAT
 CTTCCGCGCCGTGATCCACATCCCCCGCCGATCCGCCAGGGCTGGAGCGGCCCTGCTGTA
 AGATATCGGATCCTCTAGA

FIG. 51

(SEQ ID NO:64)

64 / 131

gp160.modUS4.delV1

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGTGGAGCA
 GTCTTCGTTTCGCCAACGCCACCGTGTGGGTGACCGTGTACTACGGCGTCCCCGTG
 TGGAAGGAGGCCACCACCCACCTGTTCTGCGCCAGCGACGCCAAGGCTAACAGGCCAGGC
 CCACAACTGTGTGGGCCACCCACGCCCTGCGTGTGGGCCACCGACCCCAACCCCCAGGAGGTGAACC
 TGACCAACGTGACCGAGAACCTAACATGTGGAAGAACAAACATGGTGGAGCAGATGATGAG
 GACATCATCAGCCTGTGGGACCAAGAGCCTGAAGGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG
 ACCCTGAACTGCACCGACAAGCTGGCGCCGGCGAGATCAAGAACCTGAGCTCAACAT
 CACCACCGCGTGCACGACAAGGTGAGAAGGAGTACAGCCTGTTACAAGCTGGACGTGG
 TGCCCATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCAACACCCAGCGTGTACCC
 AGGCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCG
 CCATCCTGAAGTGAAGGACAAGAACGTTAACGGCACCGGCCCTGCAAGAACCTGAGGACCC
 GTGCAGTGCACCCACGGCATCCGCCCCGTGGTGGACGACCCAGCTGCTGTGAACGGCAGCCTG
 GCCGAGGAGGAGATCGTGTGGCTCCGAGAACCTAACCGACAACGCCAACGACCATCATCGT
 GCAGCTGAACGAGTCCTGGAGATCAACTGCATCCGCCCCAACAAACACCGCTAACAGCA
 TCCACATGGCCCCGGCGCCTTACGCCACGGCGACATCATCGCGACATCCGCCCC
 CCCACTGCAACATCAGCAAGGCAACTGGACCAACACCCCTGAGCAGATCGTGGAGAACGCTG
 CGCGAGCAGTCCGCAACAACAAGACCATCATCTAACAGCAGCAGCGCCGGCGAACCTG
 GATCGTGTTCACAGCTCAACTGCGCGAGTTCTACTGCAACACCCAGCCAGCTGTT
 CAACAGCACCTGGAACATACCGAGGGAGTGAACAAGACCAAGGAGAACGACACCATCATCC
 TGCCCTGCCGATCCGCCAGATCATCAACATGTGGAGGAGTGGCAAGGCCATGTACGCC
 CCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCAATATTACCGGCCCTGCTGCTGACCCGCCAC
 GGCGGCAACAAACAACCGCACCAACGACACCGAGACCTCCGCCCGCGGCCGGAACAT
 GAAAGGACAACCTGGCGAGCGAGCTGACAAGTACAAGGTGGCGCATCGAGCCCCCTGGCG
 TGGCCCCCACCAGGCAAGCGCCGCGTGGTGCAGCGCAGAACGCGCCCGTGGCGCTGGCG
 GCCCTGTTCATCGGCTTCTGGCGCCCGGGAGCACCATGGCGCCGCTCCGTGACCCCTG
 ACCGTGAGGCCGCCAGCTGCTGAGCGCATCGTCAGCAGCAGAACAAACCTGCTGCGCGC
 CATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCC
 GCATCCTGGCGTGGAGCGTACCTGAAGGACCAAGCAGCTGCTGGCATCTGGGCTGCG
 GGCAAGCTGATCTGACCAACACCAGCTGCCCTGGAACAGCAGCTGGAGCAACAAAGAGCCTGAC
 CGAGATCTGGACAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACACCGGCC
 TGATCTACACCTGATCGAGATCGCCCAGAACAGCAGGAGAACAGCAGGAGCTGCTG
 GAGCTGGACAAGTGGGCCAGCTGTGGAACTGGTTCGACATACCAACTGGCTGTGGTACATC
 CGCATCTTCATCATGATCGTGGCGGCCCTGATCGGCTGCGCATCGTGTTCGCCGTGCTGAGC
 ATCGTAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCCTGAGACCCGCTGCCGCCAG
 CGCGGCCCGACCGCCCGAGGGCATCGAGGAGGGCGAGCGCGACCGCGACCGCA
 GCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCTGGAGCAGCTGCGCAGCCTGTGCGCTGT
 TCAGCTACCAACCGCCTGCGCACCTGCTGCTGATCGTGGCCCGATCGTGGAGCTGCTGGCC
 GCCGCGGCCCTGGAGGGCCCTGAAAGTACTGGTGGAACCTGCTGAGTACTGGAGGCCAGGAGCTG
 AAGAGCAGCGCCGTGAGCCTGTTCAACGCCACGCCATGCCGTGGCCGAGGGCACCGACCG
 CATCATCGAGATCGTGCAGCGCATCTTCCCGGCCGTGATCCACATCCCCCGCCGATCCGCCA
 GGGCCTGGAGCGGCCCTGCTGTAAGATATCGGATCCTCTAGA

FIG. 52
(SEQ ID NO:65)

65 / 131

gp160.mod.US4.delV2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGTGG
 AGCAGTCTCGTTGCCAGGCCACCGTGTGTGGTGACCGTGTACTACGGCG
 TGCCCGTGTGGAAGGAGGCCACCACCCCTGTTCTGCAGCGACGCCAAGGCTTAC
 AAGGCCAGGCCACAACGTGTGGCACCCACGCCCTGCCGCCCCACCGACCCCAACCC
 CCAGGAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGG
 TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACAGAGCCTGAAGCCCTGCGTG
 AAGCTGACCCCCCTGTGCGTACCCCTGAACTGACCGACAAGCTGACCGCAGCACCA
 CGGCACCAACAGCACCAGCGCACCAACAGCACCGGCCACAAACAGCACAGCACCA
 ACAGCACCGACAGCTGGGAGAAGATGCCGAGGGCAGATCAAGAACTGCAGCTTCAAC
 ATCGGCGCCGCCGCTGATCAACTGCAACACAGCGTGTACCCAGGCCCTGCCCAA
 GGTGAGCTCGAGCCCATTCCCCATCCACTACTGCGCCCCCGCCGGCTCGCATCCTGA
 AGTGAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAG
 TGCACCCACGGCATCCGCCGTGGTGGCAGCACCCAGCTGCTGTAACGGCAGCCTGGC
 CGAGGAGGAGATCGTGTGCGCTCCGAGAACTTACCGACAACGCCAAGACCATCATCG
 TGCAGCTGAACGAGTCGTGGAGATCAACTGCATCCGCCAAACAACAACACCGCTAAG
 AGCATCCACATCGGCCCCGGCGCGCCCTCTACGCCACGGGACATCATCGGCGACAT
 CCGCCAGGCCACTGCAACATCAGCAAGCCAACCTGGACCAACACCCCTGAGCAGATCG
 TGGAGAAGCTGCGCGAGCAGTCGGCAACAACAAGACCATCATTTCAACAGCAGCAGC
 GCGGGGAGCCCCGAGATCGTGTCCACAGCTCAACTGCGGGGAGTTCTTACTG
 CAACACCAAGCCAGCTTCAACAGCACCTGGAACATCACCAGGGAGGTGAACAAGACCA
 AGGAGAACGACACCATCATCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG
 GAGGTGGCAAGGCCATGTACGCCCCCCCCTCCGGCAGATCAAGTGCAGCAGCAA
 TATTACCGGCTGCTGCTGACCCGCGACGGCGGACCAACAACACCGCAGAACGACA
 CCGAGACCTTCGCCCCGGCGGCAACATGAAGGACAACCTGGCGAGCGAGCTGTAC
 AAGTACAAGGTGGTGCATCGAGCCCCCTGGCGTGGGCCACCCAGGCCAGCGCC
 CGTGGTGCAGCGCGAGAAGCGCGCCGTGGCGTGGGCCCTGGCGCCCTGTTCATGGCTTCTGG
 GCGCCGCGGGAGCACCATGGGCCCTCGTGAACCTGACCGTGCAGGCCAG
 CTGCTGAGCGGCATCGTCAGCAGCACCGTGTGGGCATCAAGCAGCTGCAGGCCATCCTGGCG
 GCACCTGCTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCATCCTGGCG
 TGGAGCGTACCTGAAGGACCAAGCAGCTGCTGGCATCTGGGCTGCAGCGCAAGCTG
 ATCTGCACCAACCACCGTGCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGAT
 CTGGGACAACATGACCTGGATGGAGTGGAGCGCAGATCGGCAAACACCCGGCTGA
 TCTACAACCTGATCGAGATCGCCAGAACACAGCAGGAGAACAGCAGGAGCTGCTG
 GAGCTGGACAAGTGGCCAGCCTGTGAACTGGTGTGACATCACCACCTGGCTGTGGTA
 CATCCGCATCTTCATCATGATCGTGGCGGCTGATCGGCCCTGCGCATCGTGTGCG
 TGCTGAGCATCGTAACCGCGTGCAGCCAGGGCTACAGCCCCATCAGCCTGCAGACCCGC
 CTGCCCGCCAGCGGGCCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGAGCG
 CGACCGCGACCGCAGCAACCGCCTGGTCAGGCCCTGCTGGCCCTGATCTGGACGACC
 TGCGCAGCCTGTGCCCTGTCAGCTACCCACCGCCTGCCGACCTGCTGCTGATCGTGGCC
 CGCATCGTGGAGCTGCTGGCCGCCGCGCTGGAGGGCCTGAAGTACTGGTGGAACT
 GCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAACGCCACCG
 CCATCGCCGTGGCCAGGGCACCGACCGCATCATCGAGATCGTCAGCGCAGCTTCCGC
 GCCGTGATCCACATCCCCCGCCGATCCGCCAGGGCTGGAGCGCGCCCTGCTGTAAGA
 TATCGGATCCTCTAGA

FIG. 53

(SEQ ID NO:66)

66 / 131

gp160.modUS4delV1/2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTTCGTTTCCGCCAGGCCACCGTGTGCTGTGGGTGACCGTGTACTACGGCGTCCCCTG
 TGAAGGAGGCCACCCACCACCTGTTCTGCCGCCAGCGAGGCCAAGGCTTACAAGGGCAGGC
 CCAACAACTGTGTGGCCACCCACGCCCTGCGTGTGCCCCACCGAACCCCAACCCCAAGGAGGTGAACC
 TGACCAACCGTGAACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCAATGAG
 GACATCATCAGCCTGTGGGACCAAGACCTGAAGGCCCTGCGTGGGCCGCCAGGCCCTGCC
 CAAGGTGAGCTTCGAGCCCCATCCCCATCCACTGTGCCCCCCGCCGGCTTGCCTACCTGAA
 GTGCAAGGACAAGAACGTCACGGCACGGCCCCCTGCAAGAACGTGAGCACCGTGCAGTGCA
 CCCACGGCACTCGCCCCGTGGTGGACACCCAGCTGTGTGAACGGCAACCTGGCCAGGGAG
 GAGATCGTGTGCGCTCCGAGAACTTCAACCGAACACGCCAAGACCATCATCGTGCAGCTGAA
 CGAGTCCTGTGGAGATCAACTGCATCCGCCCAACAAACACCGCTAAGAGCATCCACATCG
 GCCCCGGCCCGCCCTTACGCCACCGGCCACATCATCGGCCACATCCGCCAGGCCACTGCA
 ACATCAGCAAGGCCAACTGGACCAACACCCCTGAGCAGATCGTGGAGAACGTCGCGAGCAG
 TTCGGCAACAAACAAGACCATCATCTTCAACAGCAGCAGCGGGCGAACCCGAGATCGTGT
 CCACAGCTTCAACTGCGCGGGAGTTCTTCACTGCAACACCCAGCCAAGTGTCAACAGCAC
 CTGGAACATCACCGAGGGAGGTGAACAAGACCAAGGGAGAACGACACCATCATCCTGCCCTGCC
 GCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGCCATC
 CGCGGCCAGATCAAGTCAGCAGCAATATTACCGGCTGCTGTAACCCCGCACGGCGCAC
 CAACAAACAACCGCAACACGACACCGAGACCTTCCGCCCGGCCGGCAACATGAAGGACA
 ACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCATCGAGGCCCTGGCGTGGCCCC
 ACCCAGGCCAAGCGCCGCTGGTGCAGCGAGAACGCGCCGTTGGCGTGGCGCCCTGTT
 CATCGGCTTCTGGGCCGCCGGAGCACCATGGCGCGCTCCGTGACCCCTGACCGTGCA
 GGCCCGCCAGCTGCTGAGCGGCATCGTGCAGCAGCAGAACACCTGCTGCGGCCATCGAGG
 CCCAGCAGCACCTGCTGCACTGACCGTGTGGGCATCAAGCAGCTGCAAGGCCGATCCTG
 GCCGTGGAGCGTACCTGAAGGACCAAGCAGCTGCTGGGCATCTGGGCTGCAAGCGCAAGCT
 GATCTGCACCAACCCCGTCCCTGGAAACAGCAGCTGGAGCAACAGCAGCTGACCGAGATCT
 GGGACAAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACACCCGCTGATCTAC
 AACCTGATCGAGATCGCCCAAGAACCAGCAGGAGAACGAGCAGGAGCTGCTGGAGCTGG
 ACAAGTGGCCAGCCGTGGAACCTGGTGCACATACCAACTGGCTGTGGTACATCCGATCT
 TCATCATGATCGTGGCGCCCTGATCGGCCCTGCGCATCGTGTGCGCTGAGCATCGTGA
 ACCCGCTGCGCCAGGGCTACAGCCCCATCAGCCTGAGACCCGCTGCCGCCAGCGCGGC
 CCCGACCGCCCCGAGGGCATCGAGGAGGGCGAGCGCGACCGCGACCGCAAGCAACC
 GCCTGGTGCACGGCCCTGCTGGCCCTGATCTGGACCGACCTGCGCAGCCTGTGCGCTTCACT
 ACCACCGCTGCGCACCTGCTGCTGATCGTGGCCGCATCGTGGAGCTGCTGGCCGCCCG
 GCTGGAGGGCCCTGAAGTACTGGTGGAACTGCTGCACTGAGTACTGGAGCCAGGAGCTGAAGAGC
 AGCGCCGTGAGCCTGTTCAACGCCACGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATC
 GAGATCGTGCAGCGCATCTTCCCGCCGTGATCCACATCCCCCGCCGATCCGCCAGGGCTG
 GAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA

FIG. 54

(SEQ ID NO:67)

67 / 131

gp160.modUS4 del 128-194

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTTGCTGTGCTGCTGTGTGGAGCA
 GTCTCGTTGCCAACGCCACACCGTGTGCTGGGTGACCGTGTACTACGGCGTCCCCTG
 TGGAAAGGAGGCCACCACCCCTGTCGCCCCAGCGACGCCAAGGCTTACAAGGCCAGGC
 CCACAACTGTGGGCCACCCACGCCCTGCGTGGCCACCGACCCCAACCCCAAGGGTGAACC
 TGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCAATGAG
 GACATCATCAGCCTGTGGGACCAAGAGCCTGAAAGCCCTGCGTGAAGCTGACCCCCCTGTCGTCG
 GGGGCAGGGAACTCGCAGACCGCGTATCACCCAGGCCGCTCGCCATCCTGAAGTGCAAGGACAAGAAGTT
 CATCCCCATCCACTACTCGCCCCCCGCCGCTCGCCATCCTGAAGTGCAAGGACAAGAAGTT
 CAAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAGTCACCCACGGCATTCGCCCC
 TGGTGAGCACCCAGCTGCTGTAACGGCAGCCTGCCAGGGAGGAGATCGTGTGCGCTCC
 GAGAACTTACCGACAACGCCAAGACCATCATCGTGCAGCTGAAACGAGTCCGTGGAGATCAA
 CTGCATCCGCCCAACAAACAAACACGCGTAAGAGCATCCACATCGGCCCGGCCCTCTA
 CGCCACCGCGACATCATCGGCACATCCGCCAGGGCCACTGCAACATCAGCAAGGCCAAGT
 GGACCAACACCCCTGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCCAGCTTCAACTGCGGC
 ATCATCTTCAACAGCAGCGCGGGCGACCCCGAGATCGTGTCCACAGCTTCAACTGCGGC
 GGCAGTTCTTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAG
 GTGAACAAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCAGATCATCAA
 CATGTGGCAGGGTGGCAAGGCCATGTACGCCCTGCCATCCGCGGCCAGATCAAGTGCA
 GCAGCAATTACCGCCCTGCTGACCCCGCGACGGCGGCCACCAACAAACACCGCACCAAC
 GACACCGAGACCTTCCGCCCGCGGCCAACATGAAGGACAACTGGCGCAGCGAGCTGTA
 CAAGTACAAGTGGTGCATCGAGCCCTGGCGTGGCCCCACCCAGGCCAAGGCCCG
 TGGTGAGCGAGAAGCGCGCCGTGGGCTGGCGCCCTGTTCATCGTCTTCTGGCGCC
 CGGGAGCACCATGGCGCCGCTCCGTACCCCTGACCGTGACGGCGGCCAGCTGCTGAGC
 GGCATCGTGCAGCAGCAACACCTGCTGCGGCCATCGAGGCCAGCAGCACCTGCTGCA
 GCTGACCGTGTGGGCATCAAGCAGCTGAGGCCCGCATCTGGCCGTGGAGCAGCTG
 AGGACCAAGCAGCTGCTGGCATCTGGGCTGAGCGCGCAAGCTGATCTGCACCAACCGTG
 CCTGGAAACAGCAGCTGGAGCAACAGAGCCTGACCGAGATCTGGACAACATGACCTGGAT
 GGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCATCTACAACCTGATCGAGATCGCCC
 AGAACCCAGCAGGAGAACGAGCAGGAGCTGGAGCTGGACAAGTGGCCAAGCTG
 GAACTGGTTGACATACCAACTGGCTGTGGTACATCCGCATCTTCACTGATCGTGGCG
 CCTGATCGCCCTGCGCATCGTGTGCGCTGAGCATCGTGAACCGCGTGCAGGCCAGGGCTA
 CAGCCCCATCAGCCTGAGACCCGCTGCCAGCGCGGCCAGGCCAGGCCAGGGCA
 TCGAGGAGGAGGGCGCGAGCGCGACCGCAGCAACCGCCGTGGTGCACGGCCCTGCTG
 GCCCTGATCTGGGACGACCTGCGCAGCCCTGCGCTGCTGAGCTACCCACCGCCCTGCGCG
 CTGCTGATCGTGGCCCGCATCGTGGAGCTGCTGGGCCGCCGGCTGGGAGGCCCTGAGTAC
 TGGTGGAACTGCTGCGAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAA
 CGCCACCGCCATCGCCGTGGCCAGGGCACCGACCGCATCTGAGATCGTGCAGCGCATCTT
 CGCGCCGTGATCCACATCCCCCGCCGATCCGCCAGGGCCTGGAGCGCGCCCTGCTGTAAGA
 TATCGGATCCCTAGA

FIG. 55

(SEQ ID NO:68)

68 / 131

Env_US4_C4wt

GACACTATCATACTCCCAGCAGAATAAGACAAATTATAAACATGTGGCAAGAAGTAGG
AAAAGCAATGTATGCCCTCCCATCAGAGGACAAATTAAATGTCATCAAATATTACAG
GGCTGCTATTAACAGAGATGGTGGT

FIG. 56

(SEQ ID NO:69)

Env_SF162_C4wt

GGAACATACACTCCCAGCAGAATAAAACAAATTATAAACAGGTGGCAGGAAGTAGG
AAAAGCAATGTATGCCCTCCCATCAGAGGACAAATTAGATGTCATCAAATATTACAG
GACTGCTATTAACAAGAGATGGTGGT

FIG. 57

(SEQ ID NO:70)

Env_US4_C4mod

GACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGG
CAAGGCCATGTACGCCCTGCCAGATCAAGTGCAGCAGAACATCACCG
GCCTGCTGCTGACCCGCGACGGCGGC

FIG. 58

(SEQ ID NO:71)

Env_SF162_C4mod

GGCACCATCACCTGCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCAGGAGGTGGG
CAAGGCCATGTACGCCCTGCCAGATCCGCGGCCAGATCCGCTGCAGCAGAACATCACCG
GCCTGCTGCTGACCCGCGACGGCGGC

FIG. 59

(SEQ ID NO:72)

69 / 131

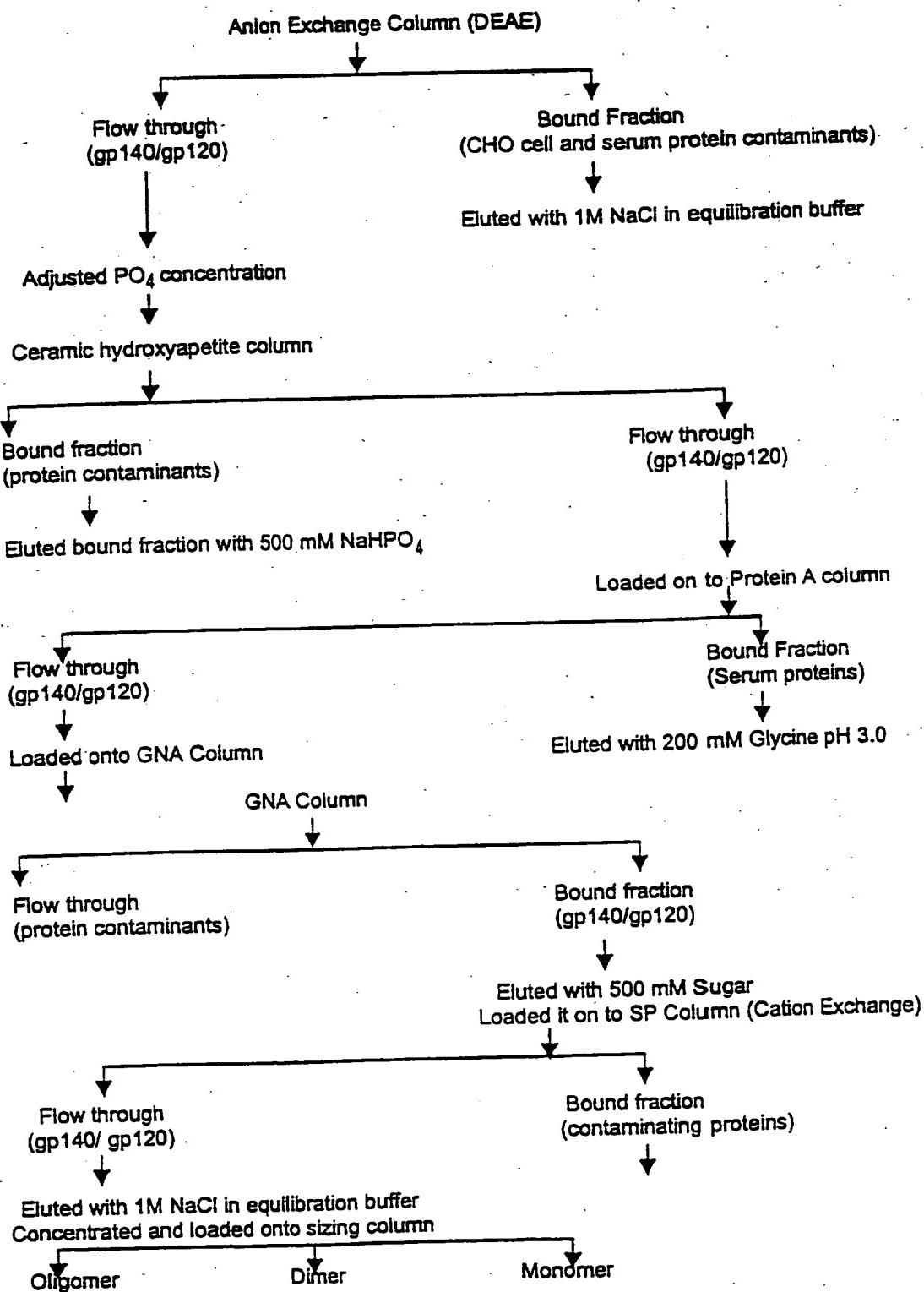


FIG. 60

70 / 131

gp160mod.us4.gag.modSF2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTTGCTGTGCTGCTGTGTTGGA
 GCAGTCTTCGTTCCGCCCAGGCCACCAACCGTGTGCTGGGTGACCGTGTACTACGGCGTG
 CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCAGCAGGCCAAGGCTTACAAG
 GCCGAGGCCACAAACGTGTGGGCCACCCACGCCCTGCGTGCCTGCCACCGACCCCCAACCCCCAG
 GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAG
 CAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGGCCCTGCGTGAAGCTG
 ACCCCCCCTGTGCGTACCCCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGGCACC
 AACAGCACCAGCGGCCACCAACAGCACCAGCGGCCACCAACAGCACCAGCACCAACAGCACC
 GACAGCTGGGAGAAGATGCCCGAGGGCAGATCAAGAACTGCAGCTTCAACATCACCACC
 AGCGTGCAGCAGAACGGTGCAGAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC
 ATCGACAAACGACAACGCCAGCTACCGCTGATCAACTGCAACACCAACAGCGTGTACCCAG
 GCCTGCCCAAGGTGAGCTTGAGCCATCCCCATCCACTACTGCGCCCGGCCGGCTTC
 GCCATCCTGAAAGTGCAAGGACAAGAAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGC
 ACCGTGCAGTGCACCCACGGCATCCGCCCGTGGTGGAGCACCCAGCTGCTGTAACGGC
 AGCCTGGCGAGGAGGAGATCGTGTGCGCTCCGAGAACTTCACCGACAACGCCAACAG
 ATCATCGTGCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCAACAAACACG
 CGTAAGAGCATCCACATCGGCCCGGCCGCTTCTACGCCACCGGCACATCATCGGC
 GACATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCCCTGAGCAG
 ATCGTGGAGAAGCTGCGCGAGCAGTCCGGCAACAAAGACCATCATCTTCAACAGCAGC
 AGCGCGGCCGACCCCGAGATCGTGTCCACAGCTTCAACTGCGCGGGCAGTTCTTCTAC
 TGCAACACCAAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGAACAAGACC
 AAGGAGAACGACACCATCATCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG
 GAGGTGGCAAGGCCATGTACGCCCGGCCAGATCAAGTGCAGCAGCAAT
 ATTACCGCCCTGCTGCTGACCCCGACGGCGGCCAACAAACAACCCACCAACGACACC
 GAGACCTTCCGCCCGGCCGGCAACATGAAGGACAACCTGGCGCAGCAGCTGTACAAG
 TACAAGGTGGTGCACATCGAGCCCTGGCGTGGCCCGACCCAGGCCAGCGCCGCGTG
 GTGCAGCGCGAGAACGCGCCGTGGCGTGGCCCTGTTCATCGGCTTCCCTGGCGCC
 GCCGGGAGCACCATGGCGCCGCGTGGCGTGGCCCTGACCCCTGACCGTGCAGGCCCGCCAGCTGCTG
 AGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGGCCATCGAGGCCAGCAGCACCTG
 CTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCGAGGCCGATCCCTGGCGTGGAGC
 TACCTGAAGGACCAGCAGCTGCTGGCATCTGGGCTGAGCGGCCAGCTGATCTGCACC
 ACCACCGTGCCTGGAACAGCAGCTGGAGCAACAAAGAGCCTGACCGAGATCTGGACAAAC
 ATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACACCGGCTGATCTACAACCTG
 ATCGAGATCGCCCAAGAACCGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAG
 TGGGCCAGCCTGGAACTGTTGACATCACAACCTGGCTGTGGTACATCCGCATCTTC
 ATCATGATCGTGGCGGCCCTGATCGGCCCTGCGCATCGTGTGCGCTGCTGAGCATCGTG
 AACCGCGTGCAGGCCAGGGCTACAGCCCATCAGCTGCGAGACCCGCCCTGCCCAGCGC
 GGCCCCGACCGCCCGAGGGCATCGAGGAGGAGGGCGGCCAGCGCAGCGCACCGCAGC
 AACCGCCTGGTGCACGCCCTGCTGGCCCTGATCTGGAGCAGCCTGCGCAGCCTGTGCTG
 TTCAGCTACCACCGCCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAGCTGCTG
 GGCGCCGCGGCTGGAGGCCCTGAAGTACTGGTGGAACTGCTGCGAGTACTGGAGGCCAG
 GAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAAGCCACCGCCATGCCGTGGCCAGGGC
 ACCGACCGCATCATCGAGATCGTGCAGCGCATCTCCGCGCCGTGATCCACATCCCCCGC
 CGCATCCGCCAGGGCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGAGAATT

FIG. 61A

(SEQ ID NO:73)

71 / 131

CGCCCCCCCCCCCCCCCCCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGC
TTGGAATAAGCCGGTGTGCGTTGTCTATATGTTATTTCAACCATATTGCCGTCTTT
GGCAATGTGAGGGCCCGAAACCTGGCCCTGTCTTGTGACGAGCATTCTAGGGGTCTT
TCCCCTCTGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTCTCTG
GAAGCTTCTTGAAGACAAACAACGTCTGAGCACCCTTGCAAGGAGCAGGGAAACCCCCA
CCTGGCGACAGGTGCCACGTTGTGAGTTGGATAGTGTGAAAGAGTCAAATGGCTCTCC
GCACAACCCAGTGCCACGTTGTGAGTTGGATAGTGTGAAAGAGTACCCATGTATGGGATCT
TCAAGCGTATTCAACAAGGGCTGAAGGATGCCAGAAGGTACCCATGTATGGGATCT
GATCTGGGCCTCGGTGCACATGCTTACATGTGTTAGTCAGGTTAAAAAACGTCTA
GGCCCCCGAACCACGGGGACGTGGTTCTTGTGAAAAACACGATAATACCATGGGCGC
CCCGGCCAGCGTGTGAGCGGGCGAGCTGGACAAGTGGAGAAGATCCGCCCTGGGCC
CGCGGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGGCCAGCCGAGCTGGAGCG
CTTCGCCGTGAACCCCGGCCCTGCTGGAGACCAGCGAGGGCTGCCAGATCTGGCCA
GCTGCAGCCCAGCCTGCAGACCGCAGCGAGGAGCTGCCAGCCTGTAACACCCGTGGC
CACCTGTACTCGTGCACCAGCGCATCGACGTCAAGGACACCAAGGAGGCCCTGGAGAA
GATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCCAGCAGGCCGCCGCCGG
CACCGGCAACAGCAGCCAGGTGAGCCAGAACCTACCCATCGTGCAGAACCTGCAGGGCA
GATGGTGACCAAGGCCATCAGCCCCCGCACCTGAAACGCCCTGGTGAAGGTGGAGGA
GAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTAGCGCCCTGAGCAGGGGCCACCC
CCAGGACCTGAACACGATGTTGAACACCGTGGGAGGCCACAGGCCCATGCAAGATGCT
GAAGGAGACCATCAACGAGGAGGCCAGATGCGCAGGCCAGTGGGACCGCGTGCACCCGG
CCCCATGCCCGGCCAGATGCGCAGGCCAGTGGGACCGCGTGCACCCGGCACACCCAG
CACCTGCAGGAGCAGATCGGCTGGATGACCAACAACCCCCCATCCCGTGGCGAGAT
CTACAAGGGTGGATCATCCTGGGCTGAACAAGATCGTGCAGGGATGTACAGCCCCACCA
CATCCTGGACATCCGCCAGGGCCCAAGGAGGCCCTCCGCGACTACGGGACCGCTCTA
CAAGACCTGCGCGTGCAGCAGGCCAGGACGTGAAGAACTGGATGACCGAGACCC
GCTGGTGAGAACGCCAACCCGACTGCAAGACCATCCTGAAGGCTCTCGGCCCCGG
CACCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGCCCCGGCACAGGCC
CGTGCTGCCAGGGCGATGAGCCAGGTGACGAACCCGGCAGCATCATGATGAGCGCG
CAACTTCGCAACCGCGGAAGACCGTCAAGTGTCTCAACTGCGGCAAGGAGGGCACAC
CGCCAGGAACCTGCCGCCCGCAAGAAGGGCTGCTGGCGTGCAGGCCAGGGCCA
CCAGATGAAGGACTGCACCGAGGCCAGGCCACCTCTGGCAAGATCTGGCCAGCTA
CAAGGGCCGCCCCGGCAACTCCTGCAGAGGCCAGGCCACCGCCCCCCCCGAGGA
GAGCTTCCGCTTCGGCGAGGAGAAGACCAACCCAGCCAGAAGCAGGAGCCATCGACAA
GGAGCTGTACCCCTGACCAAGCCTGCCAGCCTGTTGGCAACGACCCAGCAGCCAGTA
AGAATTCAAGACTCGAGCAAGTCTAGA

FIG. 61B

(SEQ ID NO:73)

72 / 131

gp160mod.SF162.gag.modSF2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGCTGTGG
 AGCAGTCTCGTTGCCAGCGCGTGGAGAACGCTGTGGGTGACCGTGTACTACGGCG
 TGCCCGTGTGGAGGAGGCCACCACCCCTGTCGCGCCAGCGACGCCAAGGCCTAC
 GACACCGAGGTGCACAACGTGTGGGCCACCCACGCCGCGTGCCTGACCGACCCCAAC
 CCAGGAGATCGTGCTGGAGAACGTGACCGAGAACCTCAACATGTGGAGAACAAACATGG
 TGGAGCAGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGGCCCTGCGTG
 AAGCTGACCCCCCTGTGCGTGAACCTGCACTGCACCAACCTGAAGAACGCCACCAACAC
 CAAGAGCAGCAACTGGAAGGAGATGGACCGCGGCGAGATCAAGAACACTGCAGCTTCAGG
 TGACCACCAAGCATCCGCAACAAGATGCAGAAGGAGTACGCCCTGTTACAAGCTGGAC
 GTGGTGCCCATCGACAAGAACACCCAGCTACAAGCTGATCAACTGCAACACCCAGCGT
 GATCACCCAGGCCCTGCCATCTGAAGTGCACAGACAAGAACGTTCAACGGCAGCGGCCCTGC
 CCGCCGGCTTCGCCATCTGAAGTGCACAGACAAGAACGTTCAACGGCAGCGGCCCTGC
 ACCAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCT
 GCTGCTGAACGGCAGCCTGGCCAGGGAGGGCGTGGTGAATCCGAGCGAGAACACTCACCG
 ACAACGCCAAGACCATCATCGCAGCTGAAGGAGAGCGTGGAGATCAACTGCAACCCCG
 CCCAACAAACACCCGCAAGAGCATCACCATCGGCCCCGGCGCGCCTTCAACGCCAC
 CGGCGACATCATCGCAGCATCCGCCAGGCCACTGCAACATCAGCGGGAGAACAGTGG
 ACAACACCCCTGAAGCAGATCGTACCAAGCTGCAGGCCAGTTCGGCAACAAGAACATC
 GTGTTCAAGCAGAGCAGCGGGCGGACCCGAGATCGTGTGACAGCTCAACTGCGG
 CGGGAGTTCTACTGCAACAGCACCCAGCTGTTCAACAGCACCTGGAACAACACCA
 TCGGCCCCAACAAACACCAACGGCACCACCATCACCTGCCCTGCCGATCAAGCAGATCATC
 AACCGCTGGCAGGAGGTGGGCAAGGCCATGTACGCCCTGCCCTGCCGAGATCCGCGGCC
 CTGCAGCAGCAACATCACCGCCCTGTCGACCCGCGACATGCGCGACAACCTGGCGAGCG
 ACACCAACCGAGATCTCCGCCCCGGCGCGGCGACATGCGCGACAACCTGGCGAGCG
 CTGTACAAGTACAAGGTGGTAAGATCGAGCCCTGGCGTGGCCCCCACCAAGGCCAA
 GCGCCGCGTGGTGCAGCGCGAGAACGCGCCGTGACCCCTGGCGCCATGTTCTGGGCT
 TCCTGGCGCCGGCGCAGCACCATGGCGCCCGAGCCTGACCGTGGCGCCATCGAGGC
 CGCCAGCTGCTGAGCGCATCGCAGCAGCGAGAACACCTGCTGCCGCCATCGAGGC
 CCAGCAGCACCTGTCGAGCTGACCGTGTGGGCATCAAGCAGCTGCGAGGCCCGCG
 TGGCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGAGCGGC
 AAGCTGATCTGCACCAACGCCGTGCCCTGGAACGCCAGCTGGAGCAACAAGAGCTGG
 CCAGATCTGGAACAAACATGACCTGGATGGAGTGGAGCGCGAGATCGACAACACCCA
 ACCTGATCTACACCCCTGATCGAGGAGAGCCAGAACCGAGCAGGAGAACGAGCAGGAG
 CTGCTGGAGCTGGACAAGTGGCCAGCCTGTTGAACTGGTTCGACATCAGCAAGTGGCT
 GTGGTACATCAAGATCTCATGATCGTGGCCCTGGCTGGCCTGCGCATCGTGT
 TCACCGCTCCCCGGCCCCCGCGGCCCGACGCCCGAGGGCATCGAGGAGGGCG
 CGAGCGCGACCGCGACCGCAGCAGGCCCTGGTGCACGCCCTGCTGCCCTGATCTGGG
 ACGACCTGCGCAGCCTGTCGCTGTTCAGCTACCAACGCCCTGCGCGACCTGATCTGATC
 GCCGCCCGCATCGTGGAGCTGCTGGGCCGCCGCGCTGGAGGCCCTGAAGTACTGGGG
 CAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGGCCGTGAGCCTGTTCGACC
 CCATCGCCATGCCGTGGCCAGGGCACCGACCGCATCATCGAGGTGGCCCAGCGCATC
 GGCGCGCCCTCCTGCACATCCCCCGCGCATCCGCCAGGGCTTCGAGCGCGCCCTGCT

FIG. 62A

(SEQ ID NO:74)

73 / 131

GTAACTCGAGCAAGTCTAGAGAATTCCGCCCCCCCCCCCCCCTCTCCCTCCCCC
CCCCCTAACGTTACTGGCCGAAGCCGCTTGGAAATAAGGCCGTGCGTTGTCTATAT
GTTATTTCCACCATATTGCCGTCTTGGCAATGTGAGGGCCGGAAACCTGGCCCTG
TCTTCTTGACGAGCATTCTAGGGTCTTCCCTCTGCCAAAGGAATGCAAGGTCTG
TTGAATGTCGTGAAGGAAGCAGTTCTCTGAAGCTTCTGAAGAACAAACAGTCTGT
AGCGACCCCTTGAGGCAGCGGAACCCCCCACCTGGCAGACAGGTGCCCTGCGGCCAA
AGCCACGGTGTATAAGATAACACCTGCAAAGCGGCACAACCCCAGTGCCACGGTGTGAGT
TGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAA
GGATGCCAGAAGGTACCCATTGTATGGGATCTGATCTGGGCCTCGGTGCACATGCT
TTACATGTGTTAGTCGAGGTTAAAAAAACGCTTAGGCCCCCGAACACGGGACGTG
GTTTCTTGAACACAGATAATACCATGGCGCCCGGCCAGCGTGTGAGCGCG
GCGAGCTGGACAAGTGGAGAAGATCCGCCTGCGCCCGCGCAAGAAGAAGTACAAG
CTGAAGCACATCGTGTGGCCAGCCGAGCTGGAGCGCTTCGCCGTGAACCCGGCCT
GCTGGAGACCGAGGGCTGCCAGATCTGGGCCAGCTGAGCCAGCTGCAGCCAGCCTGCAGA
CCGGCAGCGAGGAGCTGCCAGCCTGTACAACACCGTGGCCACCCGTACTGCGTGCAC
CAGCGCATCGACGTCAAGGACACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAA
CAAGTCCAAGAAGAAGGCCAGCAGGCCGCCGCCGGCACCGAACAGCAGCC
AGGTGAGCCAGAACTACCCATCGTCAAGAACCTGCAGGGCAGATGGTGCACCAGGCC
ATCAGCCCCCGCACCTGAACGCCCTGGTGAAGGTGGAGGAGAACGCCCTCAGCCC
CGAGGTGATCCCCATGTTAGCGCCCTGAGCGAGGGGCCACCCCCCAGGACCTGAACA
CGATGTTGAACACCGTGGCGGCCACCAGGCCCATGCAAGATGTCGAAGGAGGACATC
AACGAGGAGGCCCGAGTGGACCGCGTGCACCCCGTGCACGCCGCCCATGCC
CGGCCAGATGCCGAGCCCGGCCAGCGACATGCCGCCACCAGCACCCGTGCAGG
AGCAGATCGGCTGGATGACCAACAACCCCCCATCCCCGTGGCGAGATCTACAAGCGG
TGGATCATCCTGGCCCTGAACAAGATCGTGCAGGATGTACAGCCCCACCAGCATCCTGGA
CATCCGCCAGGGCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTACAAGACCC
TGCAGCTGAGCAGGCCAGGACGTGAAGAACTGGATGACCGAGACCCGTGTTG
CAGAACGCCAACCCGACTGCAAGACCATCCTGAAGGCTCTGCCGCCGCCGCGC
GGAGGGAGATGATGACCGCTGCCAGGGCGTGGCGGGCCACAGGCCGCGTGC
TGGCCGAGGCATGAGCCAGGTGACGAACCCGGCACCACATGATGCAAGGCCG
TTCCGCAACCAGCGGAAGACCGTCAAGTCTCAACTGCCAGGAGGGCACACCGC
CAGGAACGCCGCCGCCGCCAGAACAGGGCTGCTGGCGTGCAGGCC
AGATGAAGGACTGCACCGAGCGCCAGGCCAACCTCCTGGCAAGATCTGGCCAGCTAC
AAGGGCCGCCGGCAACTTCTGCAAGGCCGCCAGGCCACCGCCCCCGAGGA
GAGCTTCCGCTTCGGCGAGGAGAAGACCAACCCCCAGCCAGAACAGGAGGCC
AGGAGCTGTACCCCTGACCGCCTGCGCAGCCTGTTGGCAACGACCCAGCAGCCAG
TAAGAATTCAAGACTCGAGCAAGTCTAGA

FIG. 62B
(SEQ ID NO:74)

74 / 131

gp160modUS4.delV1/V2.gag.modSF2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTGTGCTGTGCTGCTGTGTTGGA
 GCAGTCTTCGTTGCCAGGCCACCACCGTGTGGGTGACCGTGACTACGGCGTG
 CCCGTGTGGAAGGAGGCCACCACCCCTGTTCTGCCAGGCCAAGGCTTACAAG
 GCCGAGGCCACAACGTGTGGGCACCCACGCCCTGCGTGCCACCGACCCAAACCCAG
 GAGGTGAACCTGACCAACGTGACCGAGAACATTCACATGTGGAAGAACAAACATGGTGGAG
 CAGATGCATGAGGACATCATCAGCCTGTGGGACAGAGCCTGAAGCCCTGCGTGGCGCC
 GGCCAGGCCCTGCCCCAAGGTGAGCTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCC
 GGCTTCGCCATCCTGAAGTGAAGGACAAGAACAGTTCAACGGCACCGGCCCTGCAAGAAC
 GTGAGCACCCTGCAGTGCACCCACGGCATCCGCCCGTGGTGAGCACCCAGCTGCTGCTG
 AACGGCAGCCTGGCCAGGAGGAGATGTCGCTGCGCTCCGAGAACATTACCGACAAACGCC
 AAGACCATCATCGTGCAGCTGAACGAGTCGTTGGAGATCAACTGCATCCGCCAACAAAC
 AACACCGCTAACAGACATCCACATCGGCCCGGCCCTTACGCCACCGGCCACATC
 ATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCAACTGGACCAACACCCCTC
 GAGCAGATCGTGGAGAACAGCTGCGCGAGCAGTCCGCAACAAACAGACCATCATTTCAAC
 AGCAGCAGCGCGCGAGATCGTGTGTTCCACAGCTTCAACTGCGCGGAGTTCAAC
 TTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCAGGCCAGGAGGTGAAC
 AAGACCAAGGAGAACGACACCATCATCTGCCCTGCCGATCCGCCAGATCATCAACATG
 TGGCAGGAGGTGGCAAGGCCATGTACGCCCTGGCCAGATCAAGTGCAGC
 AGCAATATTACCGCCCTGCTGCTGACCCCGACGGCGCACCAACAACCGCACCAAC
 GACACCGAGACCTTCCGCCCGGCCGCAACATGAAGGACAACCTGGCGCAGCGAGCTG
 TACAAGTACAAGGTGGTGCATCGAGCCCTGGCGTGGCCCTGCCAGGCAAGCGC
 CGCGTGGTGCAGCGCGAGAACGCGCCGTGGCCTGGCGCCCTGTTCATCGGCTTCTG
 GGCGCCGCCGGAGCACCATGGCGCCGCTCCGTGACCCCTGACCGTGAGGCCCGCAG
 CTGCTGAGCGCATCGTGCAGCAGAACACCTGCTGCGGCCATCGAGGCCAGCAG
 CACCTGCTGAGCTGACCGTGTGGGCATCAAGCAGCTGCAAGGCCGATCCTGGCGTG
 GAGCGCTACCTGAAGGACAGCAGCTGCTGGCATCTGGGCTGAGCGCAAGCTGATC
 TGACCAACACCCTGGAACAGCAGCTGGAGAACAGGCTGACCGAGATCTGG
 GACAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACACCGGCCTGATCTAC
 AACCTGATCGAGATCGCCAGAACCGAGCAGGAGAACGAGCAGGAGCTGCTGGAGCTG
 GACAAGTGGCCAGCCTGTTGAACTGGTTCGACATACCAACTGGCTGTGGTACATCCGC
 ATCTTCATCATGATCGTGGGCCCTGATCGGCCCTGCGCATCGTGTGCGCTGCTGAGC
 ATCGTGAACCGCGTGGCCAGGGCTACAGCCCATCAGCCTGCGAGACCGCCCTGCCCGCC
 CAGCGCGGCCCGACGCCCGAGGGCATCGAGGAGGAGGGCGAGCGCGACCGCGAC
 CGCAGCAACCGCCCTGGTGCACGGCCTGCTGGCCCTGATCTGGACGACCTGCGCAGCCTG
 TGCCTGTTAGCTACCGCCCTGCGCACCTGCTGATCGTGGCCGATCGTGGAG
 CTGCTGGGCCGCCGCGCTGGAGGGCCCTGAAGTACTGGTGAACCTGCTGCAAGTACTGG
 AGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAACGCCACGCCATGCCGTGGCC
 GAGGGCACCGACCGCATCGAGATCGTGCAGCCATCTCCGCCCGTATCCACATC
 CCCCCGCCGATCCGCCAGGGCTGGAGCGCGCCCTGCTGTAAGATACTGGATCCTCTAGA
 GAATTCCGCCCTCCCCCCCCCCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGA
 AGCCGCTTGGATAAGGCCGGTGTGCGTTGTCTATATGTTATTTCCACCATATTGCCG
 TCTTTGGCAATGTGAGGGCCGGAAACCTGGCCCTGTTCTGACGAGCATTCTAGG
 GGTCTTCCCTCTGCCAAAGGAATGCAAGGTCTGTTGAATGCGTGAAGGAAGCAGTT

FIG. 63A

(SEQ ID NO:75)

75 / 131

CCTCTGGAAGCTTCTTGAAGACAAACAGTCTGTAGCGACCCTTGCAAGGCAGCGAAC
CCCCCACCTGGCAGGGCCTCTGCGGCCAAAGCCACGTATAAGATACACCTGCA
AAGGCGGCACAACCCAGTGCACGTTGTGAGTTGGATAGTTGTGAAAGAGTCAAATGG
CTCTCCTCAAGCGTATTCAACAAGGGCTGAAGGATGCCAGAAGGTACCCATTGTATG
GGATCTGATCTGGGGCTCGGTGACATGCTTACATGTGTTAGTCGAGGTTAAAAAAA
CGTCTAGGCCCCCGAACCAACGGGACGTGGTTTCTTGAACACGATAATACCAT
GGGCGCCCGCCAGCGTGTGAGCGGGCGAGCTGGACAAGTGGGAGAAGATCCGCCT
GCGCCCGGGCAAGAAGTACAAGCTGAAGCACATCGTGTGGCCAGCCGAGCT
GGAGCGCTCGCCGTGAACCCGGCTGCTGGAGACCAGCGAGGGCTGCCAGATCCT
GGGCCAGCTGAGCCCAGCCTGCAGACCGCAGCGAGGAGCTGCGCAGCCTGTACAACAC
CGTGGCCACCCGTACTGCGTGCACCAGCGCATCGACGTCAAGGACACCAAGGAGGCC
GGAGAAGATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCCAGCAGGCCGCC
CGCCGGCACCGAACAGCAGCCAGGTGAGCCAGAACACTACCCATCGTCAGAACCTGCA
GGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCTGAACGCCCTGGTGAAGGTGG
GGAGGAGAAGGCCCTCAGCCCCGAGGTGATCCCATGTCAGGCCCTGAGCGAGGGCG
CACCCCCCAGGACCTGAACACGATGTTAACACCGTGGGCCACCAAGGCCCATGCA
GATGCTGAAGGAGACCATCAACGAGGAGGCCAGTGGGACCGCGTGCACCCGTGCA
CGCCGGCCCCATCGCCCCCGGCCAGATGCGCAGCCCCGGCAGGCACATGCCGGCAC
CACCAGCACCTGCAGGAGCAGATGGCTGGATGACCAACACCCCCCATCCCCGTGG
CGAGATCTACAAGCGGTGGATCATCTGGCCTGAACAAGATCGTGGGATGTACGCC
CACCAGCATCTGGACATCCGCAGGGCCCCAAGGAGGCCCTCCGCACTACGTGGACCG
CTTCTACAAGACCCCTGCGCGTGAACGAGGCCAGGACGTGAAGAAACTGGATGACCG
GACCCCTGCTGGTGCAGAACGCCAACCCGACTGCAAGACCATCCTGAAGGCTCTGGCC
CGCGGCCACCCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGCCCGGCCACAA
GGCCCGCGTGTGGCGAGGGATGAGCCAGGTGACGAACCCGGCGACCATCATGATGCA
GCGCGCAACTCCGAACCGAGCGGAAGACCGTCAAGTCTCAACTGCGGCAAGGAGGG
CCACACGCCAGGAACCGCCGGCTGCTGGGCTGCGCTGCCAGGGCGGCCACAA
GGGCCACCCAGATGAAGGACTGCAACGAGGCCAGGCCACTTCTGGCAAGATCTGGCC
CAGCTACAAGGGCCCGCCCCGGCAACTTCTGCAGAGCCGCCAGGCCACCGCCCC
CGAGGAGAGCTTCCGCTTGGCGAGGAGAAGACCAACCCAGGCCAGAAGCAGGAGGCC
CGACAAGGAGCTGTACCCCTGACCGCCTGCGCAGCCTGTTGGCAACGACCCAGCAG
CCAGTAAGAATTCAAGACTCGAGCAAGTCTAGA

FIG. 63B

(SEQ ID NO:75)

76 / 131

gp160.modSF162.delV2.gag.modSF2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTGTGCTGTGCTGTGTTGGAA
 GCAGTCTTCGTTGCCAGCGCGTGGAGAAGCTGTGGGTGACCGTGTACTACGGCGTG
 CCCGTGTGGAAGGAGGCCACCACCCCTGTTCTGCCAGCGACGCCAAGGCCTACGAC
 ACCGAGGTGACAACGTGTGGGCCACCCACGCCCTGCCGCCCCACGCCACCCCAACCCCCAG
 GAGATCGTGTGGAGAACGTGACCGAGAACCTCAACATGTGGAAGAACAAACATGGTGGAG
 CAGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGGCCCTGCGTGAAGCTG
 ACCCCCCCTGTGCGTACCCCTGCACTGCACCAACCTGAAGAACGCCACCAACCCAAGAGC
 AGCAACTGGAAGGAGATGGACCGCGCGAGATCAAGAACTGCAAGCTCAAGGTGGCGCC
 GGCAAGCTGATCAACTGCAACACCAGCGTGTACCCAGGCCCTGCCCAAGGTGAGCTTC
 GAGCCCATCCCCATCCACTACTGCGCCCCCGCCGCTGCCATCCTGAAGTGAACGAC
 AAGAAGTTCAACGGCAGCGGCCCTGCAACCGTGAACGGCAGCCTGGCCGAGGAGGGCGTG
 ATCCGCCCCGTGGTGAACCGCAGCTGTGTGAACGGCAGCCTGGCCGAGGAGGGCGTG
 GTGATCCCGCAGCGAGAACCTCACCGACAACGCCAAGACCATCATCGTGAAGCTGAAGGAG
 AGCGTGGAGATCAACTGCACCCGCCAACAACACACCCCGCAAGAGCATCACCATCGGC
 CCCGGCCGCCCTTCTACGCCACCGGCACATCATCGCGACATCCGCCAGGCCACTGC
 AACATCAGCGCGAGAACGATGGAACAAACACCCCTGAAGCAGATCGTGAACAGCTGAGGCC
 CAGTTCGGCAACAAGACCATCGTGTCAAGCAGAGCAGCGGGGCAACCCGAGATCGTGA
 ATGCACAGCTCAACTGCGCGGGCAGATTCTTCTACTGCAACAGCACCCAGCTGTTCAAC
 AGCACCTGGAACAAACACCATCGGCCAACAACACCAACGCCACCATCACCTGCCCTGC
 CGCATCAAGCAGATCATCAACCGCTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGGG
 ATCCGGGCCAGATCCGCTGAGCAGCAACATCACCGGCCCTGCTGACCCGCCAGGGC
 GGCAAGGAGATCAGAACACCCAGGAGATCTTCCGCCGGCGGACATGCGCGAC
 AACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCTGGCGTGGCC
 CCCACCAAGGCCAAGCGCCGCGTGGTGCAGCGCAGAAGCGCGCCGTGACCCGGCG
 ATGTTCTGGCTTCTGGGCCGCCGGCAGCACCATGGGCCCGCAGCCTGACCCCTG
 ACCGTGCAGGCCGCCAGCTGCTGAGCGGCATCGCAGCAGCAGAACACCTGCTGCGC
 GCCATCGAGGCCAGCAGCACCTGCTGAGCAGCTGAGCGCAGGAGCTGGCGTGGCG
 GCGCCGGCTGCTGGCGCTACCTGAAGGACCAGCAGCTGCTGGCGTGGCG
 TGCAGCGCAAGCTGATCTGACCAACCGCGTGGCGCAGCAGCTGGAGCAACAG
 AGCCTGGACAGATCTGGAACAAACATGACCTGGATGGAGTGGAGCGCGAGATCGACAAAC
 TACACCAACCTGATCTACACCCCTGATCGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAG
 CAGGAGCTGCTGGAGCTGGACAAGTGGCCAGCCTGTGGAACGGTCTGACATCAGCAAG
 TGGCTGTGGTACATCAAGATCTTCTCATCATGATCGTGGCGCCTGGTGGCGCTGCGCATC
 GTGTTCACCGTGTGAGCATCGTGAACCGCGTGCAGGCCAGGGCTACAGGCCCTGAGCTTC
 CAGACCCGCTTCCCCGCCCGGGCCGACCGCAGGCCAGGGCATCGAGGAGGGCG
 GGCGAGCGCAGCCGACCGCAGCAGGCCCTGGTGCAGGCCCTGCTGGCCCTGATCTGG
 GACGACCTGCGCAGCCGTGCTGCTGAGCTACCAACGCCCTGCGCAGCTGATCTGATC
 GCCGCCCGCATCGTGGAGCTGCTGGCCGCCGGCTGGAGGCCCTGAAGTACTGGGGC
 AACCTGCTGAGCTGGATCCAGGAGCTGAAGAACAGCGCCGTGAGCCTGTTGACGCC
 ATCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATGAGGTGGCCAGCGCATCGGC
 CGCGCCCTTCTGCACATCCCCGCCGATCCGCCAGGGCTCGAGCGCCCTGCTGTAA
 CTCGAGCAAGTCTAGAGAATTCCGCCCCCCCCCCCCCTCTCCCTCCCCCCCC
 TAACGTTACTGGCGAAGCCGCTTGGATAAGGCCGTGCGTTGTCTATATGTTATT
 TTCCACCATATTGCCGTCTTGGCAATGTGAGGGCCGGAAACCTGGCCCTGCTTCTT

FIG. 64A

(SEQ ID NO:76)

77 / 131

GACGAGCATTCCCTAGGGTCTTCCCTCGCAAAGGAATGCAAGGTCTGTTGAATGT
CGTGAAGGAAGCAGTCCTCTGGAAGCTCTTGAAAGACAAACACGTCGTAGCGACCC
TTGCAGGCAAGCGAACCCCCCACCTGGGACAGGGCCTCTGCGGCCAAAAGCCACGTGT
ATAAGATAACACCTGCAAAGGCGGACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGT
GGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAA
GGTACCCCCATTGTATGGATCTGATCTGGGCTCGGTGCACATGCTTACATGTGTTA
GTCGAGGTTAAAAAACGTCAGGCCCCCGAACCCACGGGACGTGGTTTCCTTGAAA
AACACGATAATACCATGGGCGCCCGCAGCGTGTGAGCGGCGGGAGCTGGACAAGT
GGGAGAAGATCCGCTGCGCCCCGGCGAAGAAGAAGTACAAGCTGAAGCACATCGTGT
GGGCCAGCCCGAGCTGGAGCGCTTCGCGGTGAACCCCGGCTGCTGGAGACCGCGAGG
GCTGCCGCCAGATCTGGGCCAGCTGCAGCCCAGCCTGCAGACCGGCGAGGAGCTGC
GCAGCTGTACAACACCGTGGCCACCGTGTACTCGCTGCACCAGCGCATCGACGTCAAGG
ACACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAAAGTCCAAGAAGAAGGCC
AGCAGGCCGCCGCCGCCGCCACCGGAAACAGCAGCCAGGTGAGCCAGAACTACCC
TCGTGCAGAACCTGCAGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCGTGAACG
CCTGGGTGAAGGTGGAGGAGAACGGCTTCAGCCCCGAGGTGATCCCCATGTTCAAGCG
CCCTGAGCGAGGGGCCACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGCGGCC
ACCAAGGCCCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCAGTGGGACC
GCGTGCACCCCGTGCACGCCGCCCCATCGCCCCCGCCAGATGCGCAGGCCCGGGCA
GCGACATGCCGGCACCAACAGCAGCCCTGCAGGAGCAGATCGGCTGGATGACCAACAACC
CCCCCATCCCCGTGGCGAGATCTACAAGCGGTGGATCATCTGGCCTGAACAAGATCG
TGGGATGTACAGCCCCACCAGCATCTGGACATCCGCCAGGGCCCCAAGGAGGCCCTCC
GGCACTACGTGGACCGCTTCTACAAGACCTGCGCCTGAGCAGGCCAGGACGTGA
AGAACTGGATGACCGAGACCCCTGCTGGTGCAGAACGCCAACCCGACTGCAAGACCATCC
TGAAGGCTCTGGCCCCGGCCACCCCTGGAGGAGATGATGACCGCCTGCCAGGGCTG
GGGGCCCCGGCAACAAGGCCCCGTGCTGGCCAGGGCGATGAGCCAGGTGACGAACCCGG
CGACCATCATGATGCGCAGCGCAACTTCCGCAACCAAGCGGAAGACCGTCAAGTGT
ACTGCGCAAGGAGGGCACACCGCCAGGAACGCGCCGGCCAGGAGACCGTCAAGTGT
GGCGCTGCCGCCGAGGGCCACAGATGAAGGACTGCACCGAGCGCAGGCCAACTTCC
TGGGCAAGATCTGGCCCAGCTACAAGGGCCGCCCGCAACTTCCCTGAGAGGCCCG
AGCCCACGCCCGAGGAGAGCTTCCGCTTGGCGAGGAGAAGACCAACCCCGAGCC
AGAACGAGGAGCCATGACAAGGAGCTGTACCCCTGACCGCCTGCCAGGCCCTG
GCAACGACCCAGCAGCCAGTAAGAAITCAGACTCGAGCAAGTCTAGA

FIG. 64B
(SEQ ID NO:76)

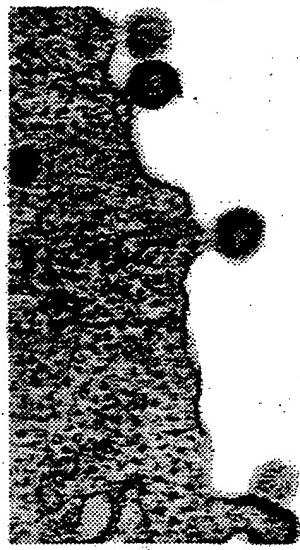


FIG. 65C



FIG. 65B



FIG. 65A

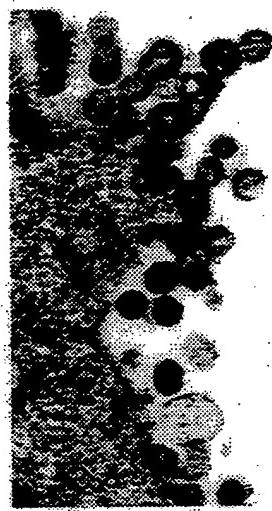


FIG. 65F



FIG. 65E



FIG. 65D

79 / 131

gp160 .modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT		
gp160 .modSF162 .delV2	(1)	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT		
gp160 .modSF162 .delV1V2	(1)	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT		
gp140 .modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT		
gp140 .mut .modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT		
gp140 .mut7 .modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT		
gp140 .mut8 .modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT		
gp120 .modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT		
Consensus	51	GAATTCGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGTGTGCT	100	
gp160 .modSF162	(51)	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG		
gp160 .modSF162 .delV2	(51)	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG		
gp160 .modSF162 .delV1V2	(51)	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG		
gp140 .modSF162	(51)	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG		
gp140 .mut .modSF162	(51)	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG		
gp140 .mut7 .modSF162	(51)	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG		
gp140 .mut8 .modSF162	(51)	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG		
gp120 .modSF162	(51)	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG		
Consensus	101	GCTGTGTGGAGCAGTCAGTCTCGTTTCGCCCAAGGCCGTGAGAAGCTGTGGG	150	
gp160 .modSF162	(101)	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG		
gp160 .modSF162 .delV2	(101)	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG		
gp160 .modSF162 .delV1V2	(101)	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG		
gp140 .modSF162	(101)	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG		
gp140 .mut .modSF162	(101)	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG		
gp140 .mut7 .modSF162	(101)	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG		
gp140 .mut8 .modSF162	(101)	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG		
gp120 .modSF162	(101)	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG		
Consensus	101	TGACCCGTGTACTACGGCGTGTGCCCGTGTGGAAAGGGCCACCAACCTG	150	

FIG. 66A-1

80 / 131

gp160 . modSF162	(151)	TTC TGG CCAGG GACGCC AAGG CCT AAC GAC ACC GAG GTGC ACA AAC GTGTG	151
gp160 . modSF162 . delV2	(151)	TTC TGG CCAGG GACGCC AAGG CCT AAC GAC ACC GAG GTGC ACA AAC GTGTG	
gp160 . modSF162 . delV1V2	(151)	TTC TGG CCAGG GACGCC AAGG CCT AAC GAC ACC GAG GTGC ACA AAC GTGTG	
gp140 . modSF162	(151)	TTC TGG CCAGG GACGCC AAGG CCT AAC GAC ACC GAG GTGC ACA AAC GTGTG	
gp140 . mut . modSF162	(151)	TTC TGG CCAGG GACGCC AAGG CCT AAC GAC ACC GAG GTGC ACA AAC GTGTG	
gp140 . mut7 . modSF162	(151)	TTC TGG CCAGG GACGCC AAGG CCT AAC GAC ACC GAG GTGC ACA AAC GTGTG	
gp140 . mut8 . modSF162	(151)	TTC TGG CCAGG GACGCC AAGG CCT AAC GAC ACC GAG GTGC ACA AAC GTGTG	
gp120 . modSF162	(151)	TTC TGG CCAGG GACGCC AAGG CCT AAC GAC ACC GAG GTGC ACA AAC GTGTG	
Consensus	201		
gp160 . modSF162	(201)	GGCC ACC CAC CGC CTGG CCC ACC GAC CCC AAC CCC CAG GAG ATCGTGC	
gp160 . modSF162 . delV2	(201)	GGCC ACC CAC CGC CTGG CCC ACC GAC CCC AAC CCC CAG GAG ATCGTGC	
gp160 . modSF162 . delV1V2	(201)	GGCC ACC CAC CGC CTGG CCC ACC GAC CCC AAC CCC CAG GAG ATCGTGC	
gp140 . modSF162	(201)	GGCC ACC CAC CGC CTGG CCC ACC GAC CCC AAC CCC CAG GAG ATCGTGC	
gp140 . mut . modSF162	(201)	GGCC ACC CAC CGC CTGG CCC ACC GAC CCC AAC CCC CAG GAG ATCGTGC	
gp140 . mut7 . modSF162	(201)	GGCC ACC CAC CGC CTGG CCC ACC GAC CCC AAC CCC CAG GAG ATCGTGC	
gp140 . mut8 . modSF162	(201)	GGCC ACC CAC CGC CTGG CCC ACC GAC CCC AAC CCC CAG GAG ATCGTGC	
gp120 . modSF162	(201)	GGCC ACC CAC CGC CTGG CCC ACC GAC CCC AAC CCC CAG GAG ATCGTGC	
Consensus	251		
gp160 . modSF162	(251)	TGG AGAAC GTG ACC CGA ACT TC AAC AT GTGG AAA ACA CAT GTGTGAG	
gp160 . modSF162 . delV2	(251)	TGG AGAAC GTG ACC CGA ACT TC AAC AT GTGG AAA ACA CAT GTGTGAG	
gp160 . modSF162 . delV1V2	(251)	TGG AGAAC GTG ACC CGA ACT TC AAC AT GTGG AAA ACA CAT GTGTGAG	
gp140 . modSF162	(251)	TGG AGAAC GTG ACC CGA ACT TC AAC AT GTGG AAA ACA CAT GTGTGAG	
gp140 . mut . modSF162	(251)	TGG AGAAC GTG ACC CGA ACT TC AAC AT GTGG AAA ACA CAT GTGTGAG	
gp140 . mut7 . modSF162	(251)	TGG AGAAC GTG ACC CGA ACT TC AAC AT GTGG AAA ACA CAT GTGTGAG	
gp140 . mut8 . modSF162	(251)	TGG AGAAC GTG ACC CGA ACT TC AAC AT GTGG AAA ACA CAT GTGTGAG	

FIG. 66A-2

81 / 131

gp120 . modSF162	(251)	TGGAGAACCGTGAACCGAGAACCTCAACATGTGGAAAGAACACATGGTGGAG	301
Consensus		TGGAGAACCGTGAACCGAGAACCTCAACATGTGGAAAGAACACATGGTGGAG	
gp160 . modSF162 . delV2	(301)	CAGATGCCACGAGGACATCATCAGGCCCTGTGGGACCAGAGCCTGAAGCCCTG	350
gp160 . modSF162 . delV1V2	(301)	CAGATGCCACGAGGACATCATCAGGCCCTGTGGGACCAGAGCCTGAAGCCCTG	
gp140 . modSF162	(301)	CAGATGCCACGAGGACATCATCAGGCCCTGTGGGACCAGAGCCTGAAGCCCTG	
gp140 . mut . modSF162	(301)	CAGATGCCACGAGGACATCATCAGGCCCTGTGGGACCAGAGCCTGAAGCCCTG	
gp140 . mut7 . modSF162	(301)	CAGATGCCACGAGGACATCATCAGGCCCTGTGGGACCAGAGCCTGAAGCCCTG	
gp140 . mut8 . modSF162	(301)	CAGATGCCACGAGGACATCATCAGGCCCTGTGGGACCAGAGCCTGAAGCCCTG	
gp120 . modSF162	(301)	CAGATGCCACGAGGACATCATCAGGCCCTGTGGGACCAGAGCCTGAAGCCCTG	
Consensus		CAGATGCCACGAGGACATCATCAGGCCCTGTGGGACCAGAGCCTGAAGCCCTG	400
gp160 . modSF162	(351)	CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCTGAAGA	450
gp160 . modSF162 . delV2	(351)	CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCAACCTGAAGA	
gp160 . modSF162 . delV1V2	(351)	CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCAACCTGAAGA	
gp140 . modSF162	(351)	CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCAACCTGAAGA	
gp140 . mut . modSF162	(351)	CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCAACCTGAAGA	
gp140 . mut7 . modSF162	(351)	CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCAACCTGAAGA	
gp140 . mut8 . modSF162	(351)	CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCAACCTGAAGA	
gp120 . modSF162	(351)	CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCAACCTGAAGA	
Consensus		CCTGAAGCTGACCCCCCTGTGGTGACCCCTGCAC TGCACTGCACCAACCTGAAGA	450
gp160 . modSF162	(401)	ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	501
gp160 . modSF162 . delV2	(401)	ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	
gp160 . modSF162 . delV1V2	(375)	ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	
gp140 . modSF162	(401)	ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	
gp140 . mut . modSF162	(401)	ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	
gp140 . mut7 . modSF162	(401)	ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	
gp140 . mut8 . modSF162	(401)	ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	
gp120 . modSF162	(401)	ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	
Consensus		ACGCCACCAACACCAAGAGCAGAACCTGGAAAGATGGACCGGGCGAG	550

FIG. 66A-3

451	gp160.modSF162 gp160.modSF162.delV2 gp160.modSF162.delV1V2 gp140.modSF162 gp140.mut.modSF162 gp140.mut7.modSF162 gp140.mut8.modSF162 gp120.modSF162 Consensus	(451) ATCAAGAACTGGCAGCTTCAGGTGACCCACCAGCATCCGCAACAAGATGCC (451) ATCAAGAACTGCAGCTTCAGGTGACCCAGCATCCGCAACAAGATGCC (376) ----- (451) ATCAAGAAACTGGCAGCTTCAGGTGACCCAGCATCCGCAACAAGATGCC (451) ATCAAGAAACTGGCAGCTTCAGGTGACCCAGCATCCGCAACAAGATGCC (451) ATCAAGAAACTGGCAGCTTCAGGTGACCCAGCATCCGCAACAAGATGCC (451) ATCAAGAAACTGGCAGCTTCAGGTGACCCAGCATCCGCAACAAGATGCC (451) ATCAAGAAACTGGCAGCTTCAGGTGACCCAGCATCCGCAACAAGATGCC 550 (451) ATCAAGAAACTGGCAGCTTCAGGTGACCCAGCATCCGCAACAAGATGCC
501	gp160.modSF162 gp160.modSF162.delV2 gp160.modSF162.delV1V2 gp140.modSF162 gp140.mut.modSF162 gp140.mut7.modSF162 gp140.mut8.modSF162 gp120.modSF162 Consensus	(501) GAAGGGAGTACGCCCTGGTTCTACAAAGCTGGACGTGGTGCCTCATCGACAAACG (478) ----- (379) ----- (501) GAAGGGAGTACGCCCTGGTTCTACAAAGCTGGACGTGGTGCCTCATCGACAAACG (501) GAAGGGAGTACGCCCTGGTTCTACAAAGCTGGACGTGGTGCCTCATCGACAAACG (501) GAAGGGAGTACGCCCTGGTTCTACAAAGCTGGACGTGGTGCCTCATCGACAAACG (501) GAAGGGAGTACGCCCTGGTTCTACAAAGCTGGACGTGGTGCCTCATCGACAAACG 551 (501) GAAGGGAGTACGCCCTGGTTCTACAAAGCTGGACGTGGTGCCTCATCGACAAACG
600	gp160.modSF162 gp160.modSF162.delV2 gp160.modSF162.delV1V2 gp140.modSF162 gp140.mut.modSF162 gp140.mut7.modSF162 gp140.mut8.modSF162 gp120.modSF162 Consensus	(551) ACAAACACCAAGCTACAAAGCTGATCAAACCTGAAACACCAGCGTGTGATCCCCAG (492) ----- (384) ----- (551) ACAAACACCAAGCTACAAAGCTGATCAAACCTGAAACACCAGCGTGTGATCCCCAG (551) ACAAACACCAAGCTACAAAGCTGATCAAACCTGAAACACCAGCGTGTGATCCCCAG

FIG. 66A-4

gp140 . mut7 . modSF162	(551)	ACPAACCCAGCTACAAGCTGATCAAACACCAGCGTGTGATCACCCAG
gp140 . mut8 . modSF162	(551)	ACPAACCCAGCTACAAGCTGATCAAACACCAGCGTGTGATCACCCAG
gp120 . modSF162	(551)	ACPAACCCAGCTACAAGCTGATCAAACACCAGCGTGTGATCACCCAG
Consensus	(551)	ACAAACCCAGCTACAAGCTGATCAAACACCAGCGTGTGATCACCCAG
	601	
gp160 . modSF162	(601)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
gp160 . modSF162 . delV2	(520)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
gp160 . modSF162 . delV1V2	(412)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
gp140 . modSF162	(601)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
gp140 . mut . modSF162	(601)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
gp140 . mut7 . modSF162	(601)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
gp140 . mut8 . modSF162	(601)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
gp120 . modSF162	(601)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
Consensus	(601)	GCCTGGCCCCAAGGTGAGCTTCGAGGCCCATCCCCATCCTCAACTACTGGCCCC
	651	
gp160 . modSF162	(651)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
gp160 . modSF162 . delV2	(570)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
gp160 . modSF162 . delV1V2	(462)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
gp140 . modSF162	(651)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
gp140 . mut . modSF162	(651)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
gp140 . mut7 . modSF162	(651)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
gp140 . mut8 . modSF162	(651)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
gp120 . modSF162	(651)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
Consensus	(651)	CGCCGGGTTTGCCTATCCTGAAAGTGCAACGCCAACGACAAGAGTTCAACGGCAGCG
	701	
gp160 . modSF162	(701)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC
gp160 . modSF162 . delV2	(620)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC
gp160 . modSF162 . delV1V2	(512)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC
gp140 . modSF162	(701)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC
gp140 . mut . modSF162	(701)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC
gp140 . mut7 . modSF162	(701)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC
gp140 . mut8 . modSF162	(701)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC
gp120 . modSF162	(701)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC
Consensus	(701)	GCCCCCTGCAACCAACCGTGAGGCCACCGTGGCAGTCGCCACGGCATCCGCC

FIG. 66A-5

84 / 131

gp160.modSF162	(751) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	900
gp160.modSF162.delV2	(670) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
gp160.modSF162.delV2	(562) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
gp160.modSF162.delV2	(751) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
gp140.modSF162	(751) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
gp140.mut.modSF162	(751) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
gp140.mut7.modSF162	(751) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
gp140.mut8.modSF162	(751) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
gp120.modSF162	(751) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
Consensus	(751) GTGGTGAGCCACCCAGCTGCTGCTGAACGGCAGGCCCTGGCCGAGGGGGGT	
	850	
gp160.modSF162	(801) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
gp160.modSF162.delV2	(720) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
gp160.modSF162.delV1V2	(612) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
gp160.modSF162.delV1V2	(801) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
gp140.modSF162	(801) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
gp140.mut.modSF162	(801) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
gp140.mut7.modSF162	(801) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
gp140.mut8.modSF162	(801) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
gp120.modSF162	(801) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
Consensus	(801) GGTGATCCGCAGCGNAGAACTTCACCGACAACGCCAAGACCATCATCGTC	
	851	
gp160.modSF162	(851) AGCTGAAGGGAGGGCTGGAGATCAACTGCAACCGCCCCAACAAACACC	
gp160.modSF162.delV2	(770) AGCTGAAGGGAGGGCTGGAGATCAACTGCAACCGCCCCAACAAACACC	
gp160.modSF162.delV1V2	(662) AGCTGAAGGGAGGGCTGGAGATCAACTGCAACCGCCCCAACAAACACC	
	900	

FIG. 66A-6

85 / 131

gp140 . modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGACCCGCCAACAAACACC	
gp140 . mut . modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGACCCGCCAACAAACACC	
gp140 . mut7 . modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGACCCGCCAACAAACACC	
gp140 . mut8 . modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGACCCGCCAACAAACACC	
gp120 . modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGACCCGCCAACAAACACC	
Consensus	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGACCCGCCAACAAACACC	950
gp160 . modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	
gp160 . modSF162 . del1V2	(820)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	
gp160 . modSF162 . del1V2	(712)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	
gp140 . modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	
gp140 . mut . modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	
gp140 . mut7 . modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	
gp140 . mut8 . modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	
gp120 . modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	
Consensus	(901)	CGCAAGAGCATCACCATCGCCCCGGCCCTTCCTAGCCACCGCGA	1000
gp160 . modSF162	(951)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	
gp160 . modSF162 . del1V2	(870)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	
gp160 . modSF162 . del1V1V2	(762)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	
gp140 . modSF162	(951)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	
gp140 . mut . modSF162	(951)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	
gp140 . mut7 . modSF162	(951)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	
gp140 . mut8 . modSF162	(951)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	
gp120 . modSF162	(951)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	
Consensus	(951)	CATCATCGGGGACATCGGCCAGGGCCACTGCAACATCAGGGCGAGAGT	

FIG. 66A-7

1050

gp160 . modSF162	(1001)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
gp160 . modSF162 . delV2	(920)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
gp160 . modSF162 . delV1V2	(812)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
gp140 . modSF162	(1001)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
gp140 . mut . modSF162	(1001)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
gp140 . mut7 . modSF162	(1001)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
gp140 . mut8 . modSF162	(1001)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
gp120 . modSF162	(1001)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
Consensus	(1001)	GGAAACAACCCCTGAAGGAGATCGTGACCAAGCTGAGGGCCAGTTGGC
	1051	
gp160 . modSF162	(1051)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
gp160 . modSF162 . delV2	(970)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
gp160 . modSF162 . delV1V2	(862)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
gp140 . modSF162	(1051)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
gp140 . mut . modSF162	(1051)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
gp140 . mut7 . modSF162	(1051)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
gp140 . mut8 . modSF162	(1051)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
gp120 . modSF162	(1051)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
Consensus	(1051)	AACAAAGACCATCGTGTTCAGCAGAGCAGGGGGGACCCCGAGATCGT
	1101	
gp160 . modSF162	(1101)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
gp160 . modSF162 . delV2	(1020)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
gp160 . modSF162 . delV1V2	(912)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
gp140 . modSF162	(1101)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
gp140 . mut . modSF162	(1101)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
gp140 . mut7 . modSF162	(1101)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
gp140 . mut8 . modSF162	(1101)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
gp120 . modSF162	(1101)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
Consensus	(1101)	GATGCCACAGCTTCAACTGCGGGGAGTTCTTCTACTGCAACAGCACCC
	1151	
gp160 . modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACACACCATTGGCCCCAACAAACCAAC
	1200	

FIG. 66A-8

gp160 . modSF162 . delV2	(1070)	AGCTGTTCAACAGCACCTGGAACAAACACCATCGGCCCAACAACACCCAAAC
gp160 . modSF162 . delV2	(962)	AGCTGTTCAACAGCACCTGGAACAAACACCATCGGCCCAACAACACCCAAAC
gp140 . modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAAACACCATCGGCCCAACAACACCCAAAC
gp140 . mut . modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAAACACCATCGGCCCAACAACACCCAAAC
gp140 . mut 7 . modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAAACACCATCGGCCCAACAACACCCAAAC
gp140 . mut 8 . modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAAACACCATCGGCCCAACAACACCCAAAC
gp120 . modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAAACACCATCGGCCCAACAACACCCAAAC
Consensus	(1151)	AGCTGTTCAACAGCACCTGGAACAAACACCATCGGCCCAACAACACCCAAAC
	1201	
gp160 . modSF162	(1201)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
gp160 . modSF162 . delV2	(1120)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
gp160 . modSF162 . delV1V2	(1012)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
gp140 . modSF162	(1201)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
gp140 . mut . modSF162	(1201)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
gp140 . mut 7 . modSF162	(1201)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
gp140 . mut 8 . modSF162	(1201)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
gp120 . modSF162	(1201)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
Consensus	(1201)	GCGACCATACCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCA
	1251	
gp160 . modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
gp160 . modSF162 . delV2	(1170)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
gp160 . modSF162 . delV1V2	(1062)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
gp140 . modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
gp140 . mut . modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
gp140 . mut 7 . modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
gp140 . mut 8 . modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
gp120 . modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
Consensus	(1251)	GGAGGTGGCAAGGCCATGTACGCCCGCCATCGCCGCCCCATCGCCGAGATCCGCT
	1300	

FIG. 66A-9

88 / 131

			1350
gp160 .modSF162	(1301)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	
gp160 .modSF162 .delV2	(1220)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	
gp160 .modSF162 .delV1V2	(1112)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	
gp140 .modSF162	(1301)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	
gp140 .mut .modSF162	(1301)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	
gp140 .mut7 .modSF162	(1301)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	
gp140 .mut8 .modSF162	(1301)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	
gp120 .modSF162	(1301)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	
Consensus	(1301)	GCAGCAGCACATCACCGGCTGCTGACCCGGACGGCAAGGAG	1400
			1351
gp160 .modSF162	(1351)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	
gp160 .modSF162 .delV2	(1270)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	
gp160 .modSF162 .delV1V2	(1162)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	
gp140 .modSF162	(1351)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	
gp140 .mut .modSF162	(1351)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	
gp140 .mut7 .modSF162	(1351)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	
gp140 .mut8 .modSF162	(1351)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	
gp120 .modSF162	(1351)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	
Consensus	(1351)	ATCAGCAACACCACCGAGATCTCCGCCCGGGCACATGCCGA	1450
			1401
gp160 .modSF162	(1401)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	
gp160 .modSF162 .delV2	(1320)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	
gp160 .modSF162 .delV1V2	(1212)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	
gp140 .modSF162	(1401)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	
gp140 .mut .modSF162	(1401)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	
gp140 .mut7 .modSF162	(1401)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	
gp140 .mut8 .modSF162	(1401)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	
gp120 .modSF162	(1401)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	
Consensus	(1401)	CAACTGGCCAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCC	

FIG. 66A-10

gp160.modsF162	(1451)	TGGCCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	1450
gp160.modSF162.delV2	(1370)	TGGCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	
gp160.modsF162.del1V1V2	(1262)	TGGCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	
gp140.modsF162	(1451)	TGGCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	
gp140.mut.modsF162	(1451)	TGGCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	
gp140.mut7.modsF162	(1451)	TGGCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	
gp140.mut8.modsF162	(1451)	TGGCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	
gp120.modsF162	(1451)	TGGCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	
Consensus	(1451)	TGGCGTGGCCCCACCAAGGCCAAGGCCCGGTTGCAGGCCGGAGAAG	
	1501		1500
gp160.modsF162	(1501)	CGGCCGTGACCTGGGGCCATGTTCTGGGCTTCCTGGGGCCGGCG	
gp160.modSF162.delV2	(1420)	CGGCCGTGACCTGGGGCCATGTTCTGGGCTTCCTGGGGCCGG	
gp160.modsF162.del1V1V2	(1312)	CGGCCGTGACCTGGGGCCATGTTCTGGGCTTCCTGGGGCCGG	
gp140.modsF162	(1501)	CGGCCGTGACCCCTGGGGCCATGTTCTGGGCTTCCTGGGGCCGG	
gp140.mut.modsF162	(1501)	AGGCCGTGACCCCTGGGGCCATGTTCTGGGCTTCCTGGGGCCGG	
gp140.mut7.modsF162	(1501)	AGGCCGTGACCCCTGGGGCCATGTTCTGGGCTTCCTGGGGCCGG	
gp140.mut8.modsF162	(1501)	AGGCCGTGACCCCTGGGGCCATGTTCTGGGCTTCCTGGGGCCGG	
gp120.modsF162	(1501)	CGC---TAACTCGAG-----	
Consensus	(1501)	CGGCCGTGACCCCTGGGGCCATGTTCTGGGCTTCCTGGGGCCGG	
	1551		1550
gp160.modsF162	(1551)	CAGCACCATGGGCCCGCAGCCTGACCCGTGACCGTGCAGGCCAGC	
gp160.modSF162.delV2	(1470)	CAGCACCATGGGCCCGCAGCCTGACCCGTGACCGTGCAGGCCAGC	
gp160.modsF162.del1V1V2	(1362)	CAGCACCATGGGCCCGCAGCCTGACCCGTGACCGTGCAGGCCAGC	
gp140.modsF162	(1551)	CAGCACCATGGGCCCGCAGCCTGACCCGTGACCGTGCAGGCCAGC	
gp140.mut.modsF162	(1551)	CAGCACCATGGGCCCGCAGCCTGACCCGTGACCGTGCAGGCCAGC	
gp140.mut7.modsF162	(1551)	CAGCACCATGGGCCCGCAGCCTGACCCGTGACCGTGCAGGCCAGC	
gp140.mut8.modsF162	(1551)	CAGCACCATGGGCCCGCAGCCTGACCCGTGACCGTGCAGGCCAGC	
gp120.modsF162	(1513)	-----	
Consensus	(1551)	CAGCACCATGGGCCCGCAGCCTGACCCGTGACCGTGCAGGCCAGC	
	1600		1600

FIG. 66A-11

90 / 131

			1650
gp160 . modSF162	(1601)	TGCTGAGGGGCATCGTCAGGCCAGAACAAACCTGGCTGGCGGCCATCGAG	
gp160 . modSF162 . delV2	(1520)	TGCTGAGGGGCATCGTCAGGCCAGAACAAACCTGGCTGGCGGCCATCGAG	
gp160 . modSF162 . delV1V2	(1412)	TGCTGAGGGGCATCGTCAGGCCAGAACAAACCTGGCTGGCGGCCATCGAG	
gp140 . modSF162	(1601)	TGCTGAGGGGCATCGTCAGGCCAGAACAAACCTGGCTGGCGGCCATCGAG	
gp140 . mut . modSF162	(1601)	TGCTGAGGGCATCGTCAGGCCAGAACAAACCTGGCTGGCGGCCATCGAG	
gp140 . mut 7 . modSF162	(1601)	TGCTGAGGGCATCGTCAGGCCAGAACAAACCTGGCTGGCGGCCATCGAG	
gp140 . mut 8 . modSF162	(1601)	TGCTGAGGGCATCGTCAGGCCAGAACAAACCTGGCTGGCGGCCATCGAG	
gp120 . modSF162	(1513)	-	
Consensus	(1601)	TGCTGAGGGCATCGTCAGGCCAGAACAAACCTGGCTGGCGGCCATCGAG	1700
gp160 . modSF162	(1651)	GCCCAGCAGCACCTGCTGCAGGTGACCCGTTGGGGCATCAAGCAGCTGCA	
gp160 . modSF162 . delV2	(1570)	GCCCAGCAGCACCTGCTGCAGGTGACCCGTTGGGGCATCAAGCAGCTGCA	
gp160 . modSF162 . delV1V2	(1462)	GCCCAGCAGCACCTGCTGCAGGTGACCCGTTGGGGCATCAAGCAGCTGCA	
gp140 . modSF162	(1651)	GCCCAGCAGCACCTGCTGCAGGTGACCCGTTGGGGCATCAAGCAGCTGCA	
gp140 . mut . modSF162	(1651)	GCCCAGCAGCACCTGCTGCAGGTGACCCGTTGGGGCATCAAGCAGCTGCA	
gp140 . mut 7 . modSF162	(1651)	GCCCAGCAGCACCTGCTGCAGGTGACCCGTTGGGGCATCAAGCAGCTGCA	
gp140 . mut 8 . modSF162	(1651)	GCCCAGCAGCACCTGCTGCAGGTGACCCGTTGGGGCATCAAGCAGCTGCA	
gp120 . modSF162	(1513)	-	
Consensus	(1651)	GCCCAGCAGCACCTGCTGCAGGTGACCCGTTGGGGCATCAAGCAGCTGCA	1750
gp160 . modSF162	(1701)	GGCCCGCGCTGGTGGCCGGCTACCTGAAGGACCCAGCTGCTGG	
gp160 . modSF162 . delV2	(1620)	GGCCCGCGCTGGTGGCCGGCTACCTGAAGGACCCAGCTGCTGG	
gp160 . modSF162 . delV1V2	(1512)	GGCCCGCGCTGGTGGCCGGCTACCTGAAGGACCCAGCTGCTGG	
gp140 . modSF162	(1701)	GGCCCGCGCTGGTGGCCGGCTACCTGAAGGACCCAGCTGCTGG	
gp140 . mut . modSF162	(1701)	GGCCCGCGCTGGTGGCCGGCTACCTGAAGGACCCAGCTGCTGG	
gp140 . mut 7 . modSF162	(1701)	GGCCCGCGCTGGTGGCCGGCTACCTGAAGGACCCAGCTGCTGG	
gp140 . mut 8 . modSF162	(1701)	GGCCCGCGCTGGTGGCCGGCTACCTGAAGGACCCAGCTGCTGG	

FIG. 66A-12

gp120 .modSF162	(1513)	GGCCCGGTTGGCCGTGGAGCTAACCTGAAGGACCAGCAGCTGCTGG	1800
Consensus	(1701)	1751	
gp160 .modSF162	(1751)	GCACTCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
gp160 .modSF162 .delV2	(1670)	GCACTCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
gp160 .modSF162 .delV1V2	(1562)	GCACTCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
gp140 .modSF162	(1751)	GCACTCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
gp140 .mut .modSF162	(1751)	GCACTCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
gp140 .mut .modSF162 .delV2	(1670)	GCACTCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
gp140 .mut 7 .modSF162	(1751)	GCACTCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
gp140 .mut 8 .modSF162	(1751)	GCACTCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
gp120 .modSF162	(1513)		
Consensus	(1751)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGACCCGCCGTGCCCTGG	
	1801	1850	
gp160 .modSF162	(1801)	AACGCCAGCTGGAGCAACAAGGGCTGGACAGATCTGGAAACACATGAC	
gp160 .modSF162 .delV2	(1720)	AACGCCAGCTGGAGCAACAAGGGCTGGACAGATCTGGAAACACATGAC	
gp160 .modSF162 .delV1V2	(1612)	AACGCCAGCTGGAGCAACAAGGGCTGGACAGATCTGGAAACACATGAC	
gp140 .modSF162	(1801)	AACGCCAGCTGGAGCAACAAGGGCTGGACAGATCTGGAAACACATGAC	
gp140 .mut .modSF162	(1801)	AACGCCAGCTGGAGCAACAAGGGCTGGACAGATCTGGAAACACATGAC	
gp140 .mut 7 .modSF162	(1801)	AACGCCAGCTGGAGCAACAAGGGCTGGACAGATCTGGAAACACATGAC	
gp140 .mut 8 .modSF162	(1801)	AACGCCAGCTGGAGCAACAAGGGCTGGACAGATCTGGAAACACATGAC	
gp120 .modSF162	(1513)		
Consensus	(1801)	AACGCCAGCTGGAGCAACAAGGGCTGGACAGATCTGGAAACACATGAC	
	1851	1900	
gp160 .modSF162	(1851)	CTGGATGGAGTGGAGATGGGGAGTCGACAACTACACCAACCTGATCTACA	
gp160 .modSF162 .delV2	(1770)	CTGGATGGAGTGGGGAGTCGACAACTACACCAACCTGATCTACA	
gp160 .modSF162 .delV1V2	(1662)	CTGGATGGAGTGGGGAGTCGACAACTACACCAACCTGATCTACA	
gp140 .modSF162	(1851)	CTGGATGGAGTGGGGAGTCGACAACTACACCAACCTGATCTACA	
gp140 .mut .modSF162	(1851)	CTGGATGGAGTGGGGAGTCGACAACTACACCAACCTGATCTACA	
gp140 .mut 7 .modSF162	(1851)	CTGGATGGAGTGGGGAGTCGACAACTACACCAACCTGATCTACA	
gp140 .mut 8 .modSF162	(1851)	CTGGATGGAGTGGGGAGTCGACAACTACACCAACCTGATCTACA	
gp120 .modSF162	(1513)		
Consensus	(1851)	CTGGATGGAGTGGGGAGTCGACAACTACACCAACCTGATCTACA	

FIG. 66A-13

92/131

		1901	1950
gp160 . modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACGAGGGAAAGAACGAGGAGGAGCTG	
gp160 . modSF162 . delV2	(1820)	CCCTGATCGAGGAGGCCAGAACGAGGGAAAGAACGAGGAGGAGCTG	
gp160 . modSF162 . delV1V2	(1712)	CCCTGATCGAGGAGGCCAGAACGAGGGAAAGAACGAGGAGGAGCTG	
gp140 . modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACGAGGGAAAGAACGAGGAGGAGCTG	
gp140 . mut . modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACGAGGGAAAGAACGAGGAGGAGCTG	
gp140 . mut7 . modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACGAGGGAAAGAACGAGGAGGAGCTG	
gp140 . mut8 . modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACGAGGGAAAGAACGAGGAGGAGCTG	
gp120 . modSF162	(1513)	-----	
Consensus	(1901)	CCCTGATCGAGGAGGCCAGAACGAGGGAAAGAACGAGGAGGAGCTG	2000
		1951	
gp160 . modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTCGACATCAGCAA	
gp160 . modSF162 . delV2	(1870)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTCGACATCAGCAA	
gp160 . modSF162 . delV1V2	(1762)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTCGACATCAGCAA	
gp140 . modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTCGACATCAGCAA	
gp140 . mut . modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTCGACATCAGCAA	
gp140 . mut7 . modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTCGACATCAGCAA	
gp140 . mut8 . modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTCGACATCAGCAA	
gp120 . modSF162	(1513)	-----	
Consensus	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTCGACATCAGCAA	2001
		2001	2050
gp160 . modSF162	(2001)	GTGGCTGTGGTACATCAAGATCTTCATCATGATCCGGCGCCTGGTGG	
gp160 . modSF162 . delV2	(1920)	GTGGCTGTGGTACATCAAGATCTTCATCATGATCCGGCGCCTGGTGG	
gp160 . modSF162 . delV1V2	(1812)	GTGGCTGTGGTACATCAAGATCTTCATCATGATCCGGCGCCTGGTGG	
gp140 . modSF162	(2001)	GTGGCTGTGGTACATCTAACCTCGAG-----	
gp140 . mut . modSF162	(2001)	GTGGCTGTGGTACATCTAACCTCGAG-----	

FIG. 66A-14

93 / 131

gp140 . mut7 . modSF162	(2001)	GTCGGCTGGTACATCTAACTCGAG	2100
gp140 . mut8 . modSF162	(2001)	GTGGCTGGTACATCTAACTCGAG	
gp120 . modSF162	(1513)	-----	
Consensus	(2001)	GTGGCTGGTACATCTAACTCGAG	
	2051		
gp160 . modSF162	(2051)	GCCTGGCATCGTGTGAGCATCGTGAACCCGGTGCAG	
gp160 . modSF162 . delV2	(1970)	GCCTGGCATCGTGTGAGCATCGTGAACCCGGTGCAG	
gp160 . modSF162 . delV2 . delV2	(1862)	GCCTGGCATCGTGTGAGCATCGTGAACCCGGTGCAG	
gp140 . modSF162	(2026)	-----	
gp140 . mut . modSF162	(2026)	-----	
gp140 . mut7 . modSF162	(2026)	-----	
gp140 . mut8 . modSF162	(2026)	-----	
gp120 . modSF162	(1513)	-----	
Consensus	(2051)		
	2101		
gp160 . modSF162	(2101)	GGCTACAGCCCCCTGAGCTCCAGACCCGGCTTCCCCGGGGGG	
gp160 . modSF162 . delV2	(2020)	GGCTACAGCCCCCTGAGCTCCAGACCCGGCTTCCCCGGGGGG	
gp160 . modSF162 . delV1V2	(1912)	GGCTACAGCCCCCTGAGCTCCAGACCCGGCTTCCCCGGGGGG	
gp140 . modSF162	(2026)	-----	
gp140 . mut . modSF162	(2026)	-----	
gp140 . mut7 . modSF162	(2026)	-----	
gp140 . mut8 . modSF162	(2026)	-----	
gp120 . modSF162	(1513)	-----	
Consensus	(2101)		
	2151		
gp160 . modSF162	(2151)	CGACCGCCCCGAGGGCATCGAGGAGGGGGGAGGGCGACCGCGACC	
gp160 . modSF162 . delV2	(2070)	CGACCGCCCCGAGGGCATCGAGGAGGGGGAGGGCGACCGCGACC	
gp160 . modSF162 . delV1V2	(1962)	CGACCGCCCCGAGGGCATCGAGGAGGGGGAGGGCGACCGCGACC	
gp140 . modSF162	(2026)	-----	
gp140 . mut . modSF162	(2026)	-----	
gp140 . mut7 . modSF162	(2026)	-----	
gp140 . mut8 . modSF162	(2026)	-----	
gp120 . modSF162	(1513)	-----	
Consensus	(2101)		
	2200		

FIG. 66A-15

94 / 131

		2250
gp160 .modSF162	(2201)	GCAGCAGCCCCCTGGTGCACGGCCTGCTGGCCCTGATCTGGGACGGACCTG
gp160 .modSF162 .delv2	(2120)	GCAGCAGCCCCCTGGTGCACGGCCTGCTGGGACGGACCTG
gp160 .modSF162 .delv1v2	(2012)	GCAGCAGCCCCCTGGTGCACGGCCTGATCTGGGACGGACCTG
gp140 .modSF162	(2026)	-----
gp140 .mut .modSF162	(2026)	-----
gp140 .mut7 .modSF162	(2026)	-----
gp140 .mut8 .modSF162	(2026)	-----
gp120 .modSF162	(1513)	-----
Consensus	(2201)	-----
		2251
gp160 .modSF162	(2251)	CGCAGCCTGGCCTCAGCTTACCAACGGCCTGGCGACCTGATCCTGAT
gp160 .modSF162 .delv2	(2170)	CGCAGCCTGGCCTACCAACGGCCTGGCGACCTGATCCTGAT
gp160 .modSF162 .delv1v2	(2062)	CGCAGCCTGGCCTACCAACGGCCTGGCGACCTGATCCTGAT
gp140 .modSF162	(2026)	-----
gp140 .mut .modSF162	(2026)	-----
gp140 .mut7 .modSF162	(2026)	-----
gp140 .mut8 .modSF162	(2026)	-----
gp120 .modSF162	(1513)	-----
Consensus	(2251)	-----
		2300
gp160 .modSF162	(2251)	CGCAGCCTGGCCTCAGCTTACCAACGGCCTGGCGACCTGATCCTGAT
gp160 .modSF162 .delv2	(2170)	CGCAGCCTGGCCTACCAACGGCCTGGCGACCTGATCCTGAT
gp160 .modSF162 .delv1v2	(2062)	CGCAGCCTGGCCTACCAACGGCCTGGCGACCTGATCCTGAT
gp140 .modSF162	(2026)	-----
gp140 .mut .modSF162	(2026)	-----
gp140 .mut7 .modSF162	(2026)	-----
gp140 .mut8 .modSF162	(2026)	-----
gp120 .modSF162	(1513)	-----
Consensus	(2251)	-----
		2301
gp160 .modSF162	(2301)	CGCGCCCCGCATCGTGGAGCTGCTGGGCCCGCGGCTGGAGGGCCCTGA
gp160 .modSF162 .delv2	(2220)	CGCGCCCCGCATCGTGGAGCTGCTGGGCCCGCGGCTGGAGGGCCCTGA
gp160 .modSF162 .delv1v2	(2112)	CGCGCCCCGCATCGTGGAGCTGCTGGGCCCGCGGCTGGAGGGCCCTGA

FIG. 66A-16

gp140.modsF162	(2026)	2400
gp140.mut.modsF162	(2026)	
gp140.mut7.modsF162	(2026)	
gp140.mut7.modsF162	(2026)	
gp140.mut8.modsF162	(2026)	
gp120.modsF162	(1513)	
Consensus	(2301)	2351
gp160.modsF162	(2351)	2401
gp160.modsF162.delV2	(2270)	
gp160.modsF162.delV1V2	(2162)	
gp140.modsF162	(2026)	
gp140.mut.modsF162	(2026)	
gp140.mut7.modsF162	(2026)	
gp140.mut8.modsF162	(2026)	
gp120.modsF162	(1513)	
Consensus	(2351)	2401
gp160.modsF162	(2401)	2450
gp160.modsF162.delV2	(2320)	
gp160.modsF162.delV1V2	(2212)	
gp140.modsF162	(2026)	
gp140.mut.modsF162	(2026)	
gp140.mut7.modsF162	(2026)	
gp140.mut8.modsF162	(2026)	
gp120.modsF162	(1513)	
Consensus	(2401)	

FIG. 66A-17

96 / 131

		2451		2500
gp160 .modSF162	(2451)	CCGCATCATCGAGGTGGCCCAGGCCATGGCCGCCCTTCCTGCACATCC		
gp160 .modSF162 .de1V2	(2370)	CCGCATCATCGAGGTGGCCCAGGCCATGGCCGCCCTTCCTGCACATCC		
gp160 .modSF162 .de1V1V2	(2262)	CCGCATCATCGAGGTGGCCCAGGCCATGGCCGCCCTTCCTGCACATCC		
gp140 .modSF162	(2026)			
gp140 .mut .modSF162	(2026)			
gp140 .mut 7 .modSF162	(2026)			
gp140 .mut 8 .modSF162	(2026)			
gp120 .modSF162	(1513)			
Consensus	(2451)			
		2501		2547
gp160 .modSF162	(2501)	CCCGCCGCCATCGGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG		
gp160 .modSF162 .de1V2	(2420)	CCCGCCGCCATCGGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG		
gp160 .modSF162 .de1V1V2	(2312)	CCCGCCGCCATCGGCCAGGGCTTCGAGGCCGCCCTGCTGTAACTCGAG		
gp140 .modSF162	(2026)			
gp140 .mut .modSF162	(2026)			
gp140 .mut 7 .modSF162	(2026)			
gp140 .mut 8 .modSF162	(2026)			
gp120 .modSF162	(1513)			
Consensus	(2501)			

FIG. 66A-18

97 / 131

Start of tPA

gp160	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	40
gp160 del V1	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
gp160 del V2	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
gp160 del V1-2	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
gp 160 del 128-194	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
gp140TM	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
gp140	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
gp140mut	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
gp120	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
Consensus	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT	
gp160	80	
gp160 del V1	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
gp160 del V2	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
gp160 del V1-2	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
gp 160 del 128-194	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
gp140TM	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
gp140	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
gp140mut	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
gp120	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
Consensus	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCCCAG	
end of tPA	120	
gp160	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp160 del V1	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp160 del V2	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp160 del V1-2	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp 160 del 128-194	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp140TM	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp140	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp140mut	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp120	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
Consensus	(81) CGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTG	
gp 160	160	
gp160 del V1	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	
gp160 del V2	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	
gp160 del V1-2	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	
gp 160 del 128-194	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	
gp140TM	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	
gp140	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	
gp140mut	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	
gp120	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	
Consensus	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA	

FIG. 66B-1

98 / 131

		161
gp160	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
gp160 del V1	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
gp160 del V2	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
gp160 del V1-2	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
gp 160 del 128-194	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
gp140TM	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
gp140	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
gp140mut	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
gp120	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
Consensus	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGGCCACAACGTGTG
		200
		201
gp160	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
gp160 del V1	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
gp160 del V2	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
gp160 del V1-2	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
gp 160 del 128-194	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
gp140TM	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
gp140	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
gp140mut	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
gp120	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
Consensus	(201)	GGCCACCCACGCCCTGCCTGCCACCGAACCCCAACCCCCAG
		240
		241
gp160	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
gp160 del V1	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
gp160 del V2	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
gp160 del V1-2	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
gp 160 del 128-194	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
gp140TM	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
gp140	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
gp140mut	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
gp120	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
Consensus	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT
		280
		281
gp160	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
gp160 del V1	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
gp160 del V2	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
gp160 del V1-2	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
gp 160 del 128-194	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
gp140TM	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
gp140	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
gp140mut	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
gp120	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
Consensus	(281)	GGAAGAACACATGGTGGAGCAGATGCATGAGGACATCAT
		320
		321
gp160	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG
gp160 del V1	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG
gp160 del V2	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGGCGCC
gp160 del V1-2	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG
gp 160 del 128-194	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG
gp140TM	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG
gp140	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG
gp140mut	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG
gp120	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG
Consensus	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCCTGAAGCTG

FIG. 66B-2

99 / 131

	361	400
gp160	(361) ACCCCCCCTGTGCGTGACCCCTGAAC TGCA CCGACA AGCTGA	
gp160 del V1	(361) ACCCCCCCTGTGCGTGACCCCTGAAC TGCA CCGACA AGCTGG	
gp160 del V2	(361) ACCCCCCCTGTGCGTGACCCCTGAAC TGCA CCGACA AGCTGA	
gp160 del V1-2	(361) GGC-----	
gp 160 del 128-194	(361) ACCCCCCCTGTGCGTGACGGGGCAGGG-----	
gp140TM	(361) ACCCCCCCTGTGCGTGACCCCTGAAC TGCA CCGACA AGCTGA	
gp140	(361) ACCCCCCCTGTGCGTGACCCCTGAAC TGCA CCGACA AGCTGA	
gp140mut	(361) ACCCCCCCTGTGCGTGACCCCTGAAC TGCA CCGACA AGCTGA	
gp120	(361) ACCCCCCCTGTGCGTGACCCCTGAAC TGCA CCGACA AGCTGA	
Consensus	(361) ACCCCCCCTGTGCGTGACCCCTGAAC TGCA CCGACA AGCTGA	
	401	440
gp160	(401) CCGGCAGCACCAACGGCACCAACAGCACCGAGCGGCCACCAA	
gp160 del V1	(401) GCGCCGGC-----	
gp160 del V2	(401) CCGGCAGCACCAACGGCACCAACAGCACCGAGCGGCCACCAA	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(401) CCGGCAGCACCAACGGCACCAACAGCACCGAGCGGCCACCAA	
gp140	(401) CCGGCAGCACCAACGGCACCAACAGCACCGAGCGGCCACCAA	
gp140mut	(401) CCGGCAGCACCAACGGCACCAACAGCACCGAGCGGCCACCAA	
gp120	(401) CCGGCAGCACCAACGGCACCAACAGCACCGAGCGGCCACCAA	
Consensus	(401) CCGGCAGCACCAACGGCACCAACAGCACCGAGCGGCCACCAA	
	441	480
gp160	(441) CAGCACCAAGCGGGCACCAACAGCACCGAGCACCAACAGCAC	
gp160 del V1	(409) -----	
gp160 del V2	(441) CAGCACCAAGCGGGCACCAACAGCACCGAGCACCAACAGCAC	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(441) CAGCACCAAGCGGGCACCAACAGCACCGAGCACCAACAGCAC	
gp140	(441) CAGCACCAAGCGGGCACCAACAGCACCGAGCACCAACAGCAC	
gp140mut	(441) CAGCACCAAGCGGGCACCAACAGCACCGAGCACCAACAGCAC	
gp120	(441) CAGCACCAAGCGGGCACCAACAGCACCGAGCACCAACAGCAC	
Consensus	(441) CAGCACCAAGCGGGCACCAACAGCACCGAGCACCAACAGCAC	
	481	520
gp160	(481) GACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAACT	
gp160 del V1	(409) -----	
gp160 del V2	(481) GACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAACT	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(481) GACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAACT	
gp140	(481) GACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAACT	
gp140mut	(481) GACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAACT	
gp120	(481) GACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAACT	
Consensus	(481) GACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAACT	
	521	560
gp160	(521) GCAGCTTCAACATCACCACCGCGTGC CGC GACA AGGTGCA	
gp160 del V1	(521) GCAGCTTCAACATCACCACCGCGTGC CGC GACA AGGTGCA	
gp160 del V2	(521) GCAGCTTCAACATCACCACCGCGTGC CGC GACA AGGTGCA	
gp160 del V1-2	(521) -----	
gp 160 del 128-194	(521) -----	
gp140TM	(521) GCAGCTTCAACATCACCACCGCGTGC CGC GACA AGGTGCA	
gp140	(521) GCAGCTTCAACATCACCACCGCGTGC CGC GACA AGGTGCA	
gp140mut	(521) GCAGCTTCAACATCACCACCGCGTGC CGC GACA AGGTGCA	
gp120	(521) GCAGCTTCAACATCACCACCGCGTGC CGC GACA AGGTGCA	
Consensus	(521) GCAGCTTCAACATCACCACCGCGTGC CGC GACA AGGTGCA	

FIG. 66B-3

100 / 131

		561
gp160	(561)	GAAGGAGTACAGCCTGTTACAAGCTGGACGTGGTGC
gp160 del V1	(465)	GAAGGAGTACAGCCTGTTACAAGCTGGACGTGGTGC
gp160 del V2	(544)	-----
gp160 del V1-2	(364)	-----
gp 160 del 128-194	(385)	-----
gp140TM	(561)	GAAGGAGTACAGCCTGTTACAAGCTGGACGTGGTGC
gp140	(561)	GAAGGAGTACAGCCTGTTACAAGCTGGACGTGGTGC
gp140mut	(561)	GAAGGAGTACAGCCTGTTACAAGCTGGACGTGGTGC
gp120	(561)	GAAGGAGTACAGCCTGTTACAAGCTGGACGTGGTGC
Consensus	(561)	GAAGGAGTACAGCCTGTTACAAGCTGGACGTGGTGC
		600
gp160	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
gp160 del V1	(505)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
gp160 del V2	(544)	-----CGCTGATCAACTGCA
gp160 del V1-2	(364)	-----
gp 160 del 128-194	(385)	-----AACTGCG
gp140TM	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
gp140	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
gp140mut	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
gp120	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
Consensus	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA
		640
gp160	(641)	ACACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
gp160 del V1	(545)	ACACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
gp160 del V2	(560)	ACACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
gp160 del V1-2	(364)	-----CAGGCCTGCCCAAGGTGAGCTT
gp 160 del 128-194	(392)	AGACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
gp140TM	(641)	ACACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
gp140	(641)	ACACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
gp140mut	(641)	ACACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
gp120	(641)	ACACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
Consensus	(641)	ACACCAGCGT GATCACCCAGGCCTGCCCAAGGTGAGCTT
		680
gp160	(681)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
gp160 del V1	(585)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
gp160 del V2	(600)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
gp160 del V1-2	(387)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
gp 160 del 128-194	(432)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
gp140TM	(681)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
gp140	(681)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
gp140mut	(681)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
gp120	(681)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
Consensus	(681)	CGAGCCC ATCCCCATCCACTACTGCGCCCCCGCCGGCTTC
		720
gp160	(721)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
gp160 del V1	(625)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
gp160 del V2	(640)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
gp160 del V1-2	(427)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
gp 160 del 128-194	(472)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
gp140TM	(721)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
gp140	(721)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
gp140mut	(721)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
gp120	(721)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG
Consensus	(721)	GCCAT CCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG

FIG. 66B-4

101 / 131

		800
gp160	(761) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
gp160 del V1	(665) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
gp160 del V2	(680) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
gp160 del V1-2	(467) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
gp 160 del 128-194	(512) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
gp140TM	(761) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
gp140	(761) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
gp140mut	(761) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
gp120	(761) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
Consensus	(761) GCCCCCTGCAAGAACGTGAGCACC GTGCAGTGCACCCACGG	
	840	
gp160	(801) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V1	(705) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V2	(720) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V1-2	(507) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp 160 del 128-194	(552) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140TM	(801) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140	(801) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140mut	(801) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp120	(801) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
Consensus	(801) CATCCGGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
	880	
gp160	(841) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGAGAACT	
gp160 del V1	(745) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
gp160 del V2	(760) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
gp160 del V1-2	(547) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
gp 160 del 128-194	(592) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
gp140TM	(841) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
gp140	(841) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
gp140mut	(841) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
gp120	(841) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
Consensus	(841) AGCCTGGCCGAGGGAGGAGATCGTGCCTCGGCTCCGAGAACT	
	920	
gp160	(881) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
gp160 del V1	(785) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
gp160 del V2	(800) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
gp160 del V1-2	(587) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
gp 160 del 128-194	(632) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
gp140TM	(881) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
gp140	(881) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
gp140mut	(881) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
gp120	(881) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
Consensus	(881) TCACCGACAACGCCAAGACC ATCATCGTCAGCTGAACGA	
	960	
gp160	(921) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
gp160 del V1	(825) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
gp160 del V2	(840) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
gp160 del V1-2	(627) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
gp 160 del 128-194	(672) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
gp140TM	(921) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
gp140	(921) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
gp140mut	(921) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
gp120	(921) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	
Consensus	(921) GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAACACG	

FIG. 66B-5

102 / 131

		1000
gp160	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
gp160 del V1	(865)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
gp160 del V2	(880)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
gp160 del V1-2	(667)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
gp 160 del 128-194	(712)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
gp140TM	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
gp140	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
gp140mut	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
gp120	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
Consensus	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG
		1040
gp160	(1001)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
gp160 del V1	(905)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
gp160 del V2	(920)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
gp160 del V1-2	(707)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
gp 160 del 128-194	(752)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
gp140TM	(1001)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
gp140	(1001)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
gp140mut.	(1001)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
gp120	(1001)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
Consensus	(1001)	CCACCGGGGACATCATCGGGACATCCGCCAGGCCACTG
		1080
gp160	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
gp160 del V1	(945)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
gp160 del V2	(960)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
gp160 del V1-2	(747)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
gp 160 del 128-194	(792)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
gp140TM	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
gp140	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
gp140mut.	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
gp120	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
Consensus	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCCTCGAGCAG
		1120
gp160	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
gp160 del V1	(985)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
gp160 del V2	(1000)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
gp160 del V1-2	(787)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
gp 160 del 128-194	(832)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
gp140TM	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
gp140	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
gp140mut.	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
gp120	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
Consensus	(1081)	ATCGTGGAGAAGCTGCGCGAGCAGTTGGCAACAACAAGA
		1160
gp160	(1121)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
gp160 del V1	(1025)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
gp160 del V2	(1040)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
gp160 del V1-2	(827)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
gp 160 del 128-194	(872)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
gp140TM	(1121)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
gp140	(1121)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
gp140mut.	(1121)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
gp120	(1121)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT
Consensus	(1121)	CCATCATCTTCAACAGCAGCGAGCGGGCGACCCCCGAGAT

FIG. 66B-6

103 / 131

			1161	1200
gp160	(1161)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
gp160 del V1	(1065)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
gp160 del V2	(1080)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
gp160 del V1-2	(867)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
gp 160 del 128-194	(912)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
gp140TM	(1161)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
gp140	(1161)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
gp140mut	(1161)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
gp120	(1161)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		
Consensus	(1161)	CGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC		1240
		1201		
gp160	(1201)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
gp160 del V1	(1105)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
gp160 del V2	(1120)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
gp160 del V1-2	(907)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
gp 160 del 128-194	(952)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
gp140TM	(1201)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
gp140	(1201)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
gp140mut	(1201)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
gp120	(1201)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		
Consensus	(1201)	TGCAACACCAGCCAGCTGTCAACAGCACCTGGAACATCA		1280
		1241		
gp160	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp160 del V1	(1145)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp160 del V2	(1160)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp160 del V1-2	(947)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp 160 del 128-194	(992)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp140TM	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp140	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp140mut	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp120	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
Consensus	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		1320
		1281		
gp160	(1281)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
gp160 del V1	(1185)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
gp160 del V2	(1200)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
gp160 del V1-2	(987)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
gp 160 del 128-194	(1032)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
gp140TM	(1281)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
gp140	(1281)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
gp140mut	(1281)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
gp120	(1281)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		
Consensus	(1281)	CCTGCCCTGCCATCCGCCAGATCATCAACATGTGGCAG		1360
		1321		
gp160	(1321)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
gp160 del V1	(1225)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
gp160 del V2	(1240)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
gp160 del V1-2	(1027)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
gp 160 del 128-194	(1072)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
gp140TM	(1321)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
gp140	(1321)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
gp140mut	(1321)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
gp120	(1321)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		
Consensus	(1321)	GAGGTGGGCAAGGCCATGTACCCCCCCCCATCCGGGCC		

FIG. 66B-7

104 / 131

			1400
gp160	(1361)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
gp160 del V1	(1265)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
gp160 del V2	(1280)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
gp160 del V1-2	(1067)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
gp 160 del 128-194	(1112)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
gp140TM	(1361)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
gp140	(1361)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
gp140mut	(1361)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
gp120	(1361)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
Consensus	(1361)	AGATCAAGTGCAGCAGCAATATTACCGCCCTGCTGCTGAC	
			1401
gp160	(1401)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
gp160 del V1	(1305)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
gp160 del V2	(1320)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
gp160 del V1-2	(1107)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
gp 160 del 128-194	(1152)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
gp140TM	(1401)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
gp140	(1401)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
gp140mut	(1401)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
gp120	(1401)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
Consensus	(1401)	CCGGCACGGCGGCCACCAACAACAAACCGCACCAACGACACC	
			1441
gp160	(1441)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
gp160 del V1	(1345)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
gp160 del V2	(1360)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
gp160 del V1-2	(1147)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
gp 160 del 128-194	(1192)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
gp140TM	(1441)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
gp140	(1441)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
gp140mut	(1441)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
gp120	(1441)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
Consensus	(1441)	GAGACCTTCCGCCCCGGCGCGCAACATGAAGGACAAC	
			1481
gp160	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
gp160 del V1	(1385)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
gp160 del V2	(1400)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
gp160 del V1-2	(1187)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
gp 160 del 128-194	(1232)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
gp140TM	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
gp140	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
gp140mut	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
gp120	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
Consensus	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGTGGTGCGCATCGA	
			1521
gp160	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp160 del V1	(1425)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp160 del V2	(1440)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp160 del V1-2	(1227)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp 160 del 128-194	(1272)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp140TM	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp140	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp140mut	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp120	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
Consensus	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	

FIG. 66B-8

105 / 131

		1600
	1561	
gp160	(1561) GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCCTGT	
gp160 del V1	(1465) GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCCTGT	
gp160 del V2	(1480) GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCCTGT	
gp160 del V1-2	(1267) GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCCTGT	
gp 160 del 128-194	(1312) GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCCTGT	
gp140TM	(1561) GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCCTGT	
gp140	(1561) GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCCTGT	
gp140mut	(1561) GTGCAGCGCGAGAAGAGCGCCGTGGGCCTGGGCCTGT	
gp120	(1561) GTGCAGCGCGAGAAGCGCTAAG-----	
Consensus	(1561) GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCCTGT	
	1601	1640
gp160	(1601) TCATCGGCTTC-CTGGGC CGCCGGAGCACCATGGCG	
gp160 del V1	(1505) TCATCGGCTTC-CTGGGC CGCCGGAGCACCATGGCG	
gp160 del V2	(1520) TCATCGGCTTC-CTGGGC CGCCGGAGCACCATGGCG	
gp160 del V1-2	(1307) TCATCGGCTTC-CTGGGC CGCCGGAGCACCATGGCG	
gp 160 del 128-194	(1352) TCATCGGCTTC-CTGGGC CGCCGGAGCACCATGGCG	
gp140TM	(1601) TCATCGGCTTC-CTGGGC CGCCGGAGCACCATGGCG	
gp140	(1601) TCATCGGCTTC-CTGGGC CGCCGGAGCACCATGGCG	
gp140mut	(1601) TCATCGGCTTC-CTGGGC CGCCGGAGCACCATGGCG	
gp120	(1583) ATATCGGATCCTCTAGA-----	
Consensus	(1601) TCATCGGCTTCNCTGGGC CGCCGGAGCACCATGGCG	
	1641	1680
gp160	(1640) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
gp160 del V1	(1544) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
gp160 del V2	(1559) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
gp160 del V1-2	(1346) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
gp 160 del 128-194	(1391) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
gp140TM	(1640) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
gp140	(1640) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
gp140mut	(1640) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
gp120	(1600) -----	
Consensus	(1641) CCGCCTCCGTGACCCGTGACCGTGCAAGGCCGCCAGCTGCT	
	1681	1720
gp160	(1680) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
gp160 del V1	(1584) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
gp160 del V2	(1599) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
gp160 del V1-2	(1386) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
gp 160 del 128-194	(1431) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
gp140TM	(1680) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
gp140	(1680) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
gp140mut	(1680) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
gp120	(1600) -----	
Consensus	(1681) GAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGCGCC	
	1721	1760
gp160	(1720) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V1	(1624) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V2	(1639) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V1-2	(1426) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp 160 del 128-194	(1471) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140TM	(1720) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140	(1720) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140mut	(1720) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp120	(1600) -----	
Consensus	(1721) ATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	

FIG. 66B-9

SUBSTITUTE SHEET (RULE 26)

106/131

			1761	1800
gp160	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
gp160 del V1	(1664)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
gp160 del V2	(1679)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
gp160 del V1-2	(1466)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
gp 160 del 128-194	(1511)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
gp140TM	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
gp140	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
gp140mut	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
gp120	(1600)	-----		
Consensus	(1761)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG		
		1801		1840
gp160	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
gp160 del V1	(1704)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
gp160 del V2	(1719)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
gp160 del V1-2	(1506)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
gp 160 del 128-194	(1551)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
gp140TM	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
gp140	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
gp140mut	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
gp120	(1600)	-----		
Consensus	(1801)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGC		
		1841		1880
gp160	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
gp160 del V1	(1744)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
gp160 del V2	(1759)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
gp160 del V1-2	(1546)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
gp 160 del 128-194	(1591)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
gp140TM	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
gp140	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
gp140mut	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
gp120	(1600)	-----		
Consensus	(1841)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCTGGAAACA		
		1881		1920
gp160	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
gp160 del V1	(1784)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
gp160 del V2	(1799)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
gp160 del V1-2	(1586)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
gp 160 del 128-194	(1631)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
gp140TM	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
gp140	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
gp140mut	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
gp120	(1600)	-----		
Consensus	(1881)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA		
		1921		1960
gp160	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		
gp160 del V1	(1824)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		
gp160 del V2	(1839)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		
gp160 del V1-2	(1626)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		
gp 160 del 128-194	(1671)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		
gp140TM	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		
gp140	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		
gp140mut	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		
gp120	(1600)	-----		
Consensus	(1921)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC		

FIG. 66B-10

SUBSTITUTE SHEET (RULE 26)

107 / 131

		1961	2000
gp160	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp160 del V1	(1864)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp160 del V2	(1879)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp160 del V1-2	(1666)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp 160 del 128-194	(1711)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp140TM	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp140	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp140mut	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp120	(1600)	-----	
Consensus	(1961)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	2040
		2001	
gp160	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V1	(1904)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V2	(1919)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V1-2	(1706)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp 160 del 128-194	(1751)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140TM	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140mut	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp120	(1600)	-----	
Consensus	(2001)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	2080
		2041	
gp160	(2040)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	
gp160 del V1	(1944)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	
gp160 del V2	(1959)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	
gp160 del V1-2	(1746)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	
gp 160 del 128-194	(1791)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	
gp140TM	(2040)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	
gp140	(2040)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	
gp140mut	(2040)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	
gp120	(1600)	-----	
Consensus	(2041)	GTGGGCCAGCTGTGAACTGGTCGACATCACCAACTGG	2120
		2081	
gp160	(2080)	CTGGTACATCCGCATCTTCATCATGATCGGGCGGCC	
gp160 del V1	(1984)	CTGGTACATCCGCATCTTCATCATGATCGGGCGGCC	
gp160 del V2	(1999)	CTGGTACATCCGCATCTTCATCATGATCGGGCGGCC	
gp160 del V1-2	(1786)	CTGGTACATCCGCATCTTCATCATGATCGGGCGGCC	
gp 160 del 128-194	(1831)	CTGGTACATCCGCATCTTCATCATGATCGGGCGGCC	
gp140TM	(2080)	CTGGTACATCCGCATCTTCATCATGATCGGGCGGCC	
gp140	(2080)	CTGGTACATC-----	
gp140mut	(2080)	CTGGTACATC-----	
gp120	(1600)	-----	
Consensus	(2081)	CTGGTACATCCGCATCTTCATCATGATCGGGCGGCC	2160
		2121	
gp160	(2120)	TGATCGGCCTGCGCATCGTGTGCGCGTGTGAGCA-----	
gp160 del V1	(2024)	TGATCGGCCTGCGCATCGTGTGCGCGTGTGAGCA-----	
gp160 del V2	(2039)	TGATCGGCCTGCGCATCGTGTGCGCGTGTGAGCA-----	
gp160 del V1-2	(1826)	TGATCGGCCTGCGCATCGTGTGCGCGTGTGAGCA-----	
gp 160 del 128-194	(1871)	TGATCGGCCTGCGCATCGTGTGCGCGTGTGAGCA-----	
gp140TM	(2120)	TGATCGGCCTGCGCATCGTGTGCGCGTGTGAGCA-----	
gp140	(2092)	-----	
gp140mut	(2092)	-----	
gp120	(1600)	-----	
Consensus	(2121)	TGATCGGCCTGCGCATCGTGTGCGCGTGTGAGCANNNN	

FIG. 66B-11

108 / 131

			2200
gp160	(2156)	- TCGTAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp160 del V1	(2060)	- TCGTAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp160 del V2	(2075)	- TCGTAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp160 del V1-2	(1862)	- TCGTAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp 160 del 128-194	(1907)	- TCGTAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	
gp140TM	(2160)	GTAAGATATCGGATCCTCTAGA-----	
gp140	(2092)	-TAAGATATCGGATCCTCTAGA-----	
gp140mut	(2092)	-TAAGATATCGGATCCTCTAGA-----	
gp120	(1600)	-----	
Consensus	(2161)	NTCGTAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	2240
		2201	
gp160	(2195)	TGCAGACCCGECTGCCGCCAGCGCGGCCGACCGCCC	
gp160 del V1	(2099)	TGCAGACCCGECTGCCGCCAGCGCGGCCGACCGCCC	
gp160 del V2	(2114)	TGCAGACCCGECTGCCGCCAGCGCGGCCGACCGCCC	
gp160 del V1-2	(1901)	TGCAGACCCGECTGCCGCCAGCGCGGCCGACCGCCC	
gp 160 del 128-194	(1946)	TGCAGACCCGECTGCCGCCAGCGCGGCCGACCGCCC	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2201)	TGCAGACCCGECTGCCGCCAGCGCGGCCGACCGCCC	2280
		2241	
gp160	(2235)	CGAGGGCATCGAGGAGGGAGGGCGGCAGCGCGGCCGAC	
gp160 del V1	(2139)	CGAGGGCATCGAGGAGGGAGGGCGGCAGCGCGACCAC	
gp160 del V2	(2154)	CGAGGGCATCGAGGAGGGAGGGCGGCAGCGCGACCAC	
gp160 del V1-2	(1941)	CGAGGGCATCGAGGAGGGAGGGCGGCAGCGCGACCAC	
gp 160 del 128-194	(1986)	CGAGGGCATCGAGGAGGGAGGGCGGCAGCGCGACCAC	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2241)	CGAGGGCATCGAGGAGGGAGGGCGGCAGCGCGACCAC	2320
		2281	
gp160	(2275)	CGCAGCAACCGCTGGTGACGGCTGCTGGCCCTGATCT	
gp160 del V1	(2179)	CGCAGCAACCGCTGGTGACGGCTGCTGGCCCTGATCT	
gp160 del V2	(2194)	CGCAGCAACCGCTGGTGACGGCTGCTGGCCCTGATCT	
gp160 del V1-2	(1981)	CGCAGCAACCGCTGGTGACGGCTGCTGGCCCTGATCT	
gp 160 del 128-194	(2026)	CGCAGCAACCGCTGGTGACGGCTGCTGGCCCTGATCT	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2281)	CGCAGCAACCGCTGGTGACGGCTGCTGGCCCTGATCT	2360
		2321	
gp160	(2315)	GGGACGACCTGCGCAGCCTGTGCCCTTTCAGCTACCACCG	
gp160 del V1	(2219)	GGGACGACCTGCGCAGCCTGTGCCCTTTCAGCTACCACCG	
gp160 del V2	(2234)	GGGACGACCTGCGCAGCCTGTGCCCTTTCAGCTACCACCG	
gp160 del V1-2	(2021)	GGGACGACCTGCGCAGCCTGTGCCCTTTCAGCTACCACCG	
gp 160 del 128-194	(2066)	GGGACGACCTGCGCAGCCTGTGCCCTTTCAGCTACCACCG	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2321)	GGGACGACCTGCGCAGCCTGTGCCCTTTCAGCTACCACCG	

FIG. 66B-12

109/131

			2400
		2361	
gp160	(2355)	CCTGGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG	
gp160 del V1	(2259)	CCTGGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG	
gp160 del V2	(2274)	CCTGGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG	
gp160 del V1-2	(2061)	CCTGGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG	
gp 160 del 128-194	(2106)	CCTGGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2361)	CCTGGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG	2440
	2401		
gp160	(2395)	CTGCTGGGCCCGCGCGGCTGGGAGGCCCTGAAGTACTGGT	
gp160 del V1	(2299)	CTGCTGGGCCCGCGCGGCTGGGAGGCCCTGAAGTACTGGT	
gp160 del V2	(2314)	CTGCTGGGCCCGCGCGGCTGGGAGGCCCTGAAGTACTGGT	
gp160 del V1-2	(2101)	CTGCTGGGCCCGCGCGGCTGGGAGGCCCTGAAGTACTGGT	
gp 160 del 128-194	(2146)	CTGCTGGGCCCGCGCGGCTGGGAGGCCCTGAAGTACTGGT	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2401)	CTGCTGGGCCCGCGCGGCTGGGAGGCCCTGAAGTACTGGT	2480
	2441		
gp160	(2435)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG	
gp160 del V1	(2339)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG	
gp160 del V2	(2354)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG	
gp160 del V1-2	(2141)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG	
gp 160 del 128-194	(2186)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2441)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG	2520
	2481		
gp160	(2475)	CGCCGTGAGCCTGTTCAACGCCACGCCATGCCGTGGCC	
gp160 del V1	(2379)	CGCCGTGAGCCTGTTCAACGCCACGCCATGCCGTGGCC	
gp160 del V2	(2394)	CGCCGTGAGCCTGTTCAACGCCACGCCATGCCGTGGCC	
gp160 del V1-2	(2181)	CGCCGTGAGCCTGTTCAACGCCACGCCATGCCGTGGCC	
gp 160 del 128-194	(2226)	CGCCGTGAGCCTGTTCAACGCCACGCCATGCCGTGGCC	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2481)	CGCCGTGAGCCTGTTCAACGCCACGCCATGCCGTGGCC	2560
	2521		
gp160	(2515)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT	
gp160 del V1	(2419)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT	
gp160 del V2	(2434)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT	
gp160 del V1-2	(2221)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT	
gp 160 del 128-194	(2266)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2521)	GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT	

FIG. 66B-13

110 / 131

FIG. 66B-14

111 / 131

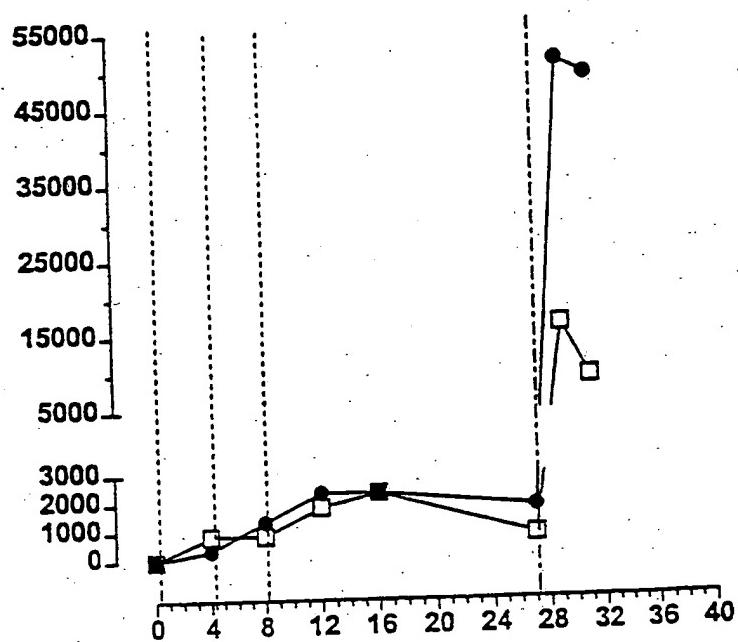


FIG. 67

112 / 131

HIV-1SF2 wt RT (PISPIET-->GIRKVL)

CCCATTAGCTTATTGAAACTGTACCACTAGAAAATTAAAGCCAGGAATGGATGGCCAAAA
 GTTAAGCAATGGCCATTGACAGAAGAAAAATAAAAGCATTAGTAGAGATATGTACAGAA
 ATGGAAAAGAAGGGAAAATTCAAAAATTGGGCTGAAAATCCATACAATACTCCAGTA
 TTTGCTATAAAGAAAAAGACAGTACTAAATGGAGAAAATAGTAGATTTCAGAGAACTT
 AATAAAAGAACTCAAGACTCTGGGAAGTTCACTAGTTAGGAATACCACACCCCGCAGGGTTA
 AAAAAGAAAAATCAGTAACTAGTATTGGATGGGTGATGCATACTTTCACTTCCCTTA
 GATAAAGACTTTAGAAAGTATACTGCATTACCATACCTAGTATAAACATGAGACACCA
 GGGATTAGATATCAGTACAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATT
 CAAAGTAGCATGACAAAATCTAGAGCCTTTAGAAAACAGAATCCAGACATAGTTATC
 TATCAAatacatggatTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAAC
 AAAATAGAGGAACGTGAGACAGCATCTGTTGAGGTGGGATTACACACCCAGACAAAAAA
 CATCAGAAAGAACCTCCATTCTTtggatggatGAACCTCCATCTGATAAATGGACA
 GTACAGCCTATAATGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAGTTA
 GTGGGAAATTGAATTGGCAAGTCAGATTATGCAGGGATTAAGTAAAGCAGTTATGT
 AAAACTCCTTAGAGGAACCAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCAGAG
 CTAGAACTGGCAGAAAACAGGGAGATTCTAAAAGAACCCAGTACATGAAGTATATTATGAC
 CCATCAAAGACTTAGTAGCAGAAATACAGAAGCAGGGCAAGGCCATGGACATATCAA
 ATTATCAAGAGCCATTAAAATCTGAAAACAGGAAAGTATGCAAGGATGAGGGTGCC
 CACACTAATGATGTAACAGTTAACAGAGGCACTAACAGGAAAGTATCCACAGAAAGCATA
 GTAATATGGGAAAGATTCTAAATTAAACTACCCATACAAAAGGAAACATGGGAGCA
 TGGTGGATGGAGTATTGGCAAGCTACCTGGATTCTGAGTGGAGTTGTCATAACCCCT
 CCCTTAGTGAATTATGGTACCACTAGCTAGGAAAGAACCCATAGTAGGAGCAGAAACTTIC
 TATGTAGATGGGCAGCTAAATAGGGAGACTAAATTAGGAAAGCAGGATATGTTACTGAC
 AGAGGAAGACAAAAGTTGTCCTCATAGCTGACACAACAAATCAGAAGACTGAATTACAA
 GCAATTCTAGCTTGCAAGGATTGGGATTAGAAGTAAACATAGTAACAGACTCACAA
 TATGCATTAGGAATCATTCAAGCACAACCAAGATAAGAGTGAATCAGAGTTAGTCAGTCAA
 ATAATAGAGCAGTTAATAAAAAGGAAAGGTCTACCTGGCATGGTACCAAGCACACAAA
 GGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGTGGAAATCAGGAAGTACTA

FIG. 68

(SEQ ID NO:77)

113 / 131

GagProtMod_SF2 (GP1)

GTCGACGCCACCATGGCGCCCGGCCAGCGTGTGAGCGCGGCGAGCTGGACAAGTGG
 GAGAAGATCCGCCTGCGCCCCGGCGAAGAAGAAGTACAAGCTGAAGCACATCGTGTGG
 GCCAGCCGGAGCTGGAGCGCTTCGCCGTGAACCCCGGCTGCTGGAGACCAGCGAGGGC
 TGCCGCCAGATCCTGGGCCAGCTGCAGCCCAGCCTGCAGACCGCAGCGAGGAGCTGCGC
 AGCCTGTACAACACCGTGGCCACCCCTGTACTGCCGTGACCGCAGTCACGTCAAGGAC
 ACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCCAG
 CAGGCCGCCGCCGCCGCCAGCGAACAGCAGCCAGGTGAGCCAGAACTACCCCATC
 GTGCAGAACCTGCAGGGCCAGATGGTGACCCAGGCCATCAGCCCCCGCACCTGAACGCC
 TGGGTGAAGGTGGTGGAGGAAGGCCCTCAGCCCCGAGGTGATCCCCATGTTAGCGCC
 CTGAGCGAGGGGCCACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGCGGCCAC
 CAGGCCGCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCGAGTGGGACCGC
 GTGCACCCCGTGCACGCCGGCCCCATCAGCCCCCGGCCAGATGCGCAGCCCCCGGCCAGC
 GACATCGCCGGCACACCAGCACCCCTGCAGGAGCAGATGGCTGGATGACCAACACCC
 CCCATCCCCGTGGCGAGATCTACAAGCGGTGGATCATCCTGGCCTGAACAAGATCGTG
 CGGATGTACAGCCCCACCAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCCG
 GACTACGTGGACCCTCTACAAGACCCCTGCCGCTGAGCAGGCCAGGACGTGAAG
 AACTGGATGACCGAGACCCCTGCTGGTGCAAGCACAACCCGACTGCAAGACCATCCTG
 AAGGCTCTCGGCCCCGGCCACCCCTGGAGGAGATGATGACCGCCTGCCAGGGCTGG
 GGCCCCGGCCACAAGGCCCGGTGCTGGCCGAGGCGATGAGCCAGGTGACGAACCCGGCG
 ACCATCATGATGCAGCGGGCAACTCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAAC
 TCGGGCAAGGAGGGCACACCGCCAGGAACCTGCCGCCGGCAAGAAGGGCTGCTGG
 CGCTGCCGCCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTA
 GGGAAAGATCTGGCCTTCTACAAGGAAGGCCAGGAATTTCTCAGAGCAGACCAGAG
 CCAACAGCCCCACCAGAACAGAGACTCAGGTTGGGAGGAGAAACAACCTCCCTCTCAG
 AAGCAGGAGCCGATAGACAAGGAACGTATCCTTAACTCCCTCAGATCACTCTTGGC
 AACGACCCCTCGTCACAGTAAGGATCGCGGCCAGCTCAAGGAGGCCGCTGCTGACACCG
 GCGCCGACGACACCGTGTGGAGGAGATGAAACCTGCCGGCAAGTGGAGCCAAAGATGA
 TCGGGGGATCGGGGCTTCATCAAGGTGCCAGTACGACCAAGATCCCCGTGGAGATCT
 GCGGCCACAAGGCCATGGCACCGTGTGGTGGGCCACCCCGTGAACATCATCGGCC
 GCAACCTGCTGACCCAGATCGGCTGCACCCCTGAACCTCCCCATCAGCCCCATCGAGACGG
 TGCCCGTGAAGCTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCTGTAAG
 AATTC

FIG. 69

(SEQ ID NO:78)

114 / 131

GagProtMod.SF2 (GP2)

GTCGACGCCACCATGGCGCCCGCGCCAGCGTGCTGAGCGGGCGAGCTGGACAAGTGG
 GAGAAGATCCGCTCGCCCCGGCGGAAGAAGAAGTACAAGCTGAAGCACATCGTGTGG
 GCCAGCCGCGAGCTGGAGCGCTCGCCGTGAACCCGGCTGCTGGAGACCAGCGAGGGC
 TGCCGCCAGATCCTGGCCAGCTGCAGCCCAGCCTGCAGACCGCAGCGAGGAGCTGCGC
 AGCCTGTACAACACCGTGGCCACCCCTGACTGCGTGCACCAGCGATCGACGTCAAGGAC
 ACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCCAG
 CAGGCCGCCGCCGCCGCCAGCAGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATC
 GTGCAGAACCTGCAGGGCCAGATGGTGCACCAGGCCATCAGCCCCGACCCCTGAACGCC
 TGGGTGAAGGTGGTGGAGGAGAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTAGCGCC
 CTGAGCGAGGGGCCACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGCCGGCCAC
 CAGGCCGCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCAGTGGAGCCAGC
 GTGCACCCCGTGCACGCCGCCCATGCCCGGCCAGATGCGCAGGCCAGGCCAGC
 GACATGCCGGCACCAACCAGCACCTGCAGGAGCAGATCGCTGGATGACCAACAACCC
 CCCATCCCGTGGCGAGATCTACAAGCGGTGGATCATCTGGCCTGAACAAAGATCGT
 CGGATGTACAGCCCCACCAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCCGC
 GACTACGTGGACCGCTTCTACAAGACCCCTGCGCGCTGAGCAGGCCAGCCAGGTGAAG
 AACTGGATGACCGAGACCCCTGCTGGTGCAGAACGCCAACCCGACTGCAAGACCATCCTG
 AAGGCTCTGGCCCCGGCCACCCCTGGAGGAGATGATGACCGCTGCCAGGGCGTGGG
 GGCCCCGGCCACAAGGCCCGGTGCTGGCGAGGCATGAGCCAGGTGACGAACCCGGCG
 ACCATCATGATGCAGCGGCAACTCCGCAACCAGCGGAAGACCGTCAAGTGTCAAC
 TCGGGCAAGGAGGGCACACCGCCAGGAACCTGCGCGCCAGGGCAAGAAGGCTGCTGG
 CGCTGGCCCCGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGTAATTTTA
 GGGAAAGATCTGGCTTCCTACAAGGGAAGGCCAGGGATTTCAGAGCAGACCAGAG
 CCAACAGCCCCACCAAGAGAGCTCAGGTTGGGGAGGAGAAAACAACCTCCCTCTCAG
 AACCAGGAGCCGATAGACAAGGAACGTATCCTTAACTCCCTCAGATCACTCTTGGC
 AACGACCCCTCGTACAGTAAGGATGGGGGGCAACTCAAGGAAGCGCTGCTGATAACAG
 GAGCAGATGATAACAGTATTAGAAGAAATGAATTGCCAGGAAATGAAACCAAAATGA
 TAGGGGGATGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACTGTAGAAATCT
 GTGGACATAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAA
 GAAATCTGTTGACCCAGATCGGCTGCACCTTGAACCTCCCATCAGCCCTATTGAGACGG
 TGCCCGTGAAGTTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAATGCCATTGTAAG
 AATTC

FIG. 70

(SEQ ID NO:79)

115 / 131

FS(+)_ProInact_RTopt_YM

GC GG CG CG AAGG AC ACCA AA TG AA AG ATT G C ACT G AG AG AC AGG CT AA TTTT TAGG GA
AG AT CT GG CCTT CCT AC AAGG GA AGG C AGGG A ATT T CT CAG AG CAG ACC AGG CAA
CAG CCC C ACC AGA AG AG AG C TT CAG GTT GGG GAGG AG AAA CA ACT CC CT CAG AAG C
AGG AG CC GAT AG AC AAGG AACT GT AT C TT TA ACT CC CT CAG AT C ACT C TT GG C AAC G
AC CC CT CG T CACA AT AAGG AT CG GGG GCA ACT CAAGG AAG CG CT G CT G AT AC AGG AG C
AG AT G AT A CAG T ATT AG AAG AA AT GA ATT GC CAGG AAA AT GG AA ACC AAA AT GA TAGG
GGG GAT CG GGG GCTT CAT CAAG GT GAGG CAGT AC GACC AG AT AC CT GT TAG AA AT CT GT GG
AC AT AA AG CT AT AGGT AC AGT ATT AGT TAGG AC CT AC AC CT GT CA AC AT A ATT GG AAG AAA
TCT GTT GAC CC AG AT CGG CT G CAC CT TG AACT TCCC AT CAG CC CT ATT GAG AC GG GT GCC
CGT GAAG TT GAAG CC GGG AT GG AC GG CCA AG GT CAAG CA AT GG C ATT GAC CG GAGGA
GAAG AT CAAGG CC CT GG TG AG AT CT G CAC CG AG AT GG AGA AGG GAG GG CA AG AT CAG CAA
GAT CG G C C C GAG A ACC C TACA AC ACC C C C G T T C G C AT CAAG A AGA AGG AC AG CAC
CAAG TGG CG CAAG CT GG TG ACT TCC CG GAG CT GA AC AAG CG C ACC C AGG ACT T CT GG GA
GGT GCAG CT GG GCA T C C C C A C C C G C G G C T GA AG A AGA AGA AG GCG T GAC CG T GCT
GG AC GT GG GCG AC G C CT ACT T CAG CG T G C C C T GG AC A AGG ACT T C G C AAG T AC ACC G C
CTT C ACC AT C C C C A G C AT CA AC AC GAG ACC C C C G G C AT C C G C T ACC AGT AC AC G T G C T
GCC C CAG GG CT GG AAG GG CAG C C C G C C AT CT T C C AG AG CAG C AT G ACC A AG AT C C T G G A
GCC C TT C C G CA AG C AG A ACC C C G AC AT CG T G AT CT ACC AGG C C C C C T G T AC GT GGG C AG
CG AC CT GG AG AT CG G C C AG C ACC G C ACC A AG AT CG AGG AG CT G C G C C AG C ACC T G C T G C G
CT TGG G C T T C ACC ACC C C C G AC A AG A AG C ACC A AG A AGG AG C C C C C T C T G G AT G G G
CT AC GAG CT G C ACC C C G AC A AG T GG ACC G T G CAG C C C AT C AT G CT G C C G AG A AGG AC AG
CT TGG ACC G T G A AC G AC AT CC C AG A AG G C T G G G C A AG G C T G A AC T G G G C C AG C AG AT C T A
CG C C G G C AT CA AG GT GA AG C AG C T G T G C A AG C T G C T G C G G C ACC A AG G C C T G ACC G A
GG T G AT C C C C T G ACC G AG G G C C G AG C T G G AG C T G G C C G AG A ACC G C G AG AT C C T G A A
GG AG C C C G T G C AC G AG G T G T A C T AC G ACC C C AG C A AG G AC C T G G T G G C C G AG AT C C A G A A
G C A G G G C C A G G G C C A G T G G A C C T A C C A G A T C T A C C A G G A G C C C T T C A A G A A C C T G A A G A C
C G G C A A G T A C G C C C G A T G C G C G G C C C A C A C C A A C G A C G T G A A G C A G C T G A C C G A G G C
C G T G C A G A A G G T G A G C A C C G A G A G C A T C G T G A T C T G G G C A A G A T C C C A A G T T C A A G C T

FIG. 71A

(SEQ ID NO:80)

116 / 131

GCCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGAT
CCCCGAGTGGAGTCGTAAACACCCCCCCCCTGGTGAAGCTGTGGTACCAAGCTGGAGAA
GGAGGCCATCGTGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAA
GCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATGCCGA
CACCAACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCT
GGAGGTGAACATCGTACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGGCCGA
CAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGT
GTACCTGGCCTGGGTGCCGCCACAAGGCATCGCGGCAACGAGCAGGTGGACAAGCT
GGTAGCGCCGGCATCCGAAGGTGCTGTTCTGAACGGCATCGATGGCGGCATCGTGT
CTACCAAGTACATGGACGACCTGTACGTGGCAGCGCGGCCCTAGGATCGATTAAAAGCT
TCCCGGGGCTAGCACCGGTGAATT

FIG. 71B

(SEQ ID NO:80)

117 / 131

FS(+)_ProtInact_RTopt_YMWM

GCGGCCGCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTTAGGGA
AGATCTGGCCTTCCATAAGGGAAGGCCAGGGATTTCAGAGCAGACCAGAGCCAA
CAGCCCCACCAGAAGAGAGCTTCAGGTTGGGGAGGAGAAAACAACCTCCCTCTCAGAAC
AGGAGCCGATAGACAAGGAACGTATCCTTAACCTCCCTCAGATCACTCTTGGCAACG
ACCCCTCGTCACAATAAGGATCGGGGGCACTCAAGGAAGCGCTGCTCGATAACAGGAGC
AGATGATAACAGTATTAGAAGAAATGAATTGCCAGGAAAATGAAACCAAAATGATAGG
GGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACTGTAGAAATCTGTGG
ACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAA
TCTGTTGACCCAGATCGGCTGCACCTTGAACCTCCCCTCAGCCCTATTGAGACGGTGCC
CGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCATTGACCGAGGA
GAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAA
GATCGGCCCCGAGAACCCCTACAACACCCCCGTGTCGCCATCAAGAAGAAGGACAGCAC
CAAGTGGCGCAAGCTGGTGGACTTCCCGAGCTGAACAAGCGCACCCAGGACTCTGGGA
GGTGCAGCTGGCATCCCCCACCCGCCGGCTGAAGAAGAAGAGCGTGAACCGTGCTG
GGACGTGGCGACGCCCTACCTCAGCGTGCCTGGACAAGGACTTCCGCAAGTACACCGC
CTTCACCATCCCCAGCATCAACACGAGACCCCCGGCATCCGCTACCACTACAACGTGCT
GCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTCCAGAGCAGCATGACCAAGATCCTGG
GCCCTTCCGCAAGCAGAACCCCGACATCGTGAACCTACAGGCCCCCTGTACGTGGCAG
CGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCCAGCACCTGCTGCG
CTGGGGCTTCACCACCCCCGACAAGAACGACCAAGAGGAGCCCCCTTCCGCCATCGA
GCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCAGAACGGACAGCTGGAC
CGTGAACGACATCCAGAACGACTGGTGGCAAGCTGAACCTGGCCAGCCAGATCTACGCC
CATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCCGGCACCAAGGCCCTGACCGAGGTGAT
CCCCCTGACCGAGGAGGCCAGCTGGAGCTGGCGAGAACCGCGAGATCCTGAAGGAGGC
CGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCAGATCCAGAACGG
CCAGGGCCAGTGGACCTACCAGATCTACAGGAGCCCTCAAGAACCTGAAGAACCGGCAA
GTACGCCCGCATGCCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCA
GAAGGTGAGCACCAGAGCATCGTGAACCTGGGCAAGATCCCCAAGTTCAAGCTGCCAT

FIG. 72A

(SEQ ID NO:81)

118 / 131

CCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCGA
GTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCC
CATCGTGGCGCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGCTGGG
CAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATGCCGACACCAC
CAACCAAGAACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGT
GAACATCGTGAACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCCCACAAGAG
CGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCT
GGCCTGGGTGCCGCCACAAGGGCATGGCGCAACGAGCAGGTGGACAAGCTGGTGAG
CGCCGGCATCCGAAGGTGCTGTTCTGAACGGCATCGATGGCGCATCGTGAICTACCA
GTACATGGACGACCTGTACGTGGCAGCGGGGCCCTAGGATCGATTAAAAGCTTCCC GG
GGCTAGCACCGGTGAATTC

FIG. 72B

(SEQ ID NO:81).

119 / 131

FS(-)_ProtMod_RTopt_YM

GCGGCCGCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTCTTCCGCC
 AGGACCTGGCTTCTGCAGGGCAAGGCCGAGTTCAGCAGCGAGCACCCGCGCCA
 ACAGCCCCACCCGCGCGAGCTGAGGTGTGGGCGGCGAGAACAAACAGCCTGAGCGAGG
 CCGCGCCGACCGCCAGGGCACCGTGAGCTCAACTCCCCAGATCACCTGTGGCAGC
 GCCCCCTGGTACCATCAGGATCGCGGCCAGCTCAAGGAGGCGCTGCTGACACCGCG
 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGCAAGTGGAAAGCCAAAGATGATCG
 GCGGGATCGGGGCTTCATCAAGGTGCGGCACTACGACCAGATCCCCGTGGAGATCTGCG
 GCCACAAGGCCATCGGCACCGTGCTGGTGGCCCCACCCCGTGAACATCATCGGCCGCA
 ACCTGCTGACCCAGATCGGCTGCACCCCTGAACCTCCCCATCAGCCCCATCGAGACGGTGC
 CCGTGAAGCTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCTGACCGAGG
 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA
 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTCGCCATCAAGAAGAAGGACAGCA
 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG
 AGGTGCAGCTGGCATCCCCACCCCGCCGGCTGAAGAAGAAGAAGACCGTGAACGTGC
 TGGACGTGGCGACGCCCTACTTCAGCGTCCCCCTGGACAAGGACTTCCGCAAGTACACCG
 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGATCCGCTACCAAGTACAACGTGC
 TGCCCCAGGGCTGGAAGGGAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG
 AGCCCTTCCGCAAGCAGAACCCGACATCGTGAATCTACCAAGGCCCCCTGTACGTGGCA
 GCGACCTGGAGATCGGCCAGCACCGACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC
 GCTGGGGCTTCACCACCCCGACAAGAACGACCAAGAAGGAGCCCCCTTCTGTGGATGG
 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAAGCTGCTGCCGAGAACGGACATCT
 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGCCAGCCAGATCT
 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCCGAGAACGGCCCTGACCG
 AGGTGATCCCCCTGACCGAGGAGGCCAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA
 AGGAGCCCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGA
 AGCAGGGCCAGGGCCAGTGGACCTACCAAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGA
 CGCGCAAGTACGCCGCATGCCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGG
 CGTGAGAAGGTGAGCACCGAGAGCATCGTGAATCTGGGGCAAGATCCCCAAGTCAAGC

FIG. 73A

(SEQ ID NO:82)

120 / 131

TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA
TCCCCGAGTGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA
AGGAGCCCATCGTGGCGCCGAGACCTTCTACGTGGACGGCGCCGCAACCGCGAGACCA
AGCTGGGAAGGCCGGCTACGTGACCGACCGGGCCGGCAGAAGGTGGTGAGCATGCCG
ACACCAACCAACCAGAACGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC
TGGAGGTGAACATCGTACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCCCC
ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG
TGTACCTGGCCTGGGTGCCGCCACAAGGCATCGCGCAACGAGCAGGTGGACAAGC
TGGTGAGCGCCGGCATCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCATCGTA
TCTACCAGTACATGGACGACCTGTACGTGGCAGCGCGGCCCTAGGATCGATTAAAAGC
TTCCCGGGGCTAGCACCGGTGAATT

FIG. 73B

(SEQ ID NO:82)

121 / 131

FS(-)_ProtMod_RTopt_YMWM

GCGGCCGCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTCTTCCGCC
 AGGACCTGGCCTTCCTGCAGGGCAAGGCCGCGAGTTCAAGCAGCGAGCAGACCCGCCA
 ACAGCCCCACCCGCCGAGCTGCAGGTGTGGGGCGCGAGAACAAACAGCCTGAGCGAGG
 CCGCGCCGACCGCCAGGGCACCGTGAGCTCAACTCCCCAGATCACCTGTGGCAGC
 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCCTGCTGACACCGCG
 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCGGCAAGTGGAAAGCCAAGATGATCG
 GCGGGATCGGGGCTTCATCAAGGTGGCGCAGTAGCACCAGATCCCGTGGAGATCTGCG
 GCCACAAGGCCATCGGACCGTGCTGGGGCCCACCCCGTGAACATCATCGGCCGCA
 ACCTGCTACCCAGATCGGCTGCACCTGAACCTCCCCATCAGCCCCATCGAGACCGGTGC
 CCGTGAAGCTGAAGCCGGGATGGACGGCCCAAGGTCAAGCAGTGGCCCTGACCGAGG
 AGAAGATCAAGGCCCTGGTGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA
 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTGCCATCAAGAAGAAGGACAGCA
 CCAAGTGGCGCAAGCTGGTGACTTCCCGAGCTGAACAAGCGCACCCAGGACTCTGGG
 AGGTGCAGCTGGGATCCCCACCCCGCCGGCTGAAGAAGAAGAAGAGCGTGACCGTGC
 TGGACGTGGCGACGCCTACTTCAGCGTGGCCATCAAGGACTTCCGAAGTACACCG
 CCTTCACCATCCCCAGCATCAACACCGAGACCCCCGGCATCCGCTACCGTACAACGTGC
 TGCCCCAGGGCTGGAAGGGAGCCCCGCATCTCCAGAGCAGCATGACCAAGATCTGG
 AGCCCTTCCGAAGCAGAACCCGACATCGTATCTACAGGCCCCCTGTACGTGGCA
 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCCAGCACCTGCTGC
 GCTGGGCTTCACCAACCCCCGACAAGAACGACCAAGAAGGAGCCCCCTTCCGCCATCG
 AGCTGCACCCGACAAGTGGACCGTGAGCCATCATGCTGCCAGAAGGACAGCTGGA
 CCGTGAACGACATCCAGAAGCTGGTGAGCTGAAGCTGGCCAGCCAGATCTACGCCG
 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCCGGCACCAAGGCCCTGACCGAGGTGA
 TCCCCCTGACCGAGGAGGCCAGCTGGAGCTGGCCAGAACCGCAGATCTGAAGGAGC
 CCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGAGATCCAGAAGCAGG
 GCCAGGGCCAGTGGACCTACCAGATCTACAGGAGCCCTCAAGAACCTGAAGGACCGCG
 AGTACGCCCGCATGCCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC
 AGAAGGTGAGCAGGAGAGCATCGTATCTGGGCAAGATCCCCAAGTCAAGCTGCCCA

FIG. 74A

(SEQ ID NO:83)

122 / 131

TCCAGAAGGAGACCTGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG
AGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAAGCTGGAGAAGGAGC
CCATCGTGGCGCCGAGACCTCTACGTGGACGGCGCCAACCGCGAGACCAAGCTGG
GCAAGGCCGGCTACGTGACCGACCAGGGCCGGCAGAAGGTGGTGAGCATGCCGACACCA
CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG
TGAACATCGTACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGGCCGACAAGA
GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC
TGGCCTGGTGCCCCCCCACAAGGGCATCGGCGCAACGAGCAGGTGGACAAGCTGGTGA
GCGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGATCGTGTACCTACC
AGTACATGGACGACCTGTACGTGGCAGCGGGGCCCTAGGATCGATTAAAAGCTTCCCG
GGGCTAGCACCGGTGAATTC

FIG. 74B

(SEQ ID NO:83)

123 / 131

FS(-)_ProtMod_RTopt(+)

GCGGCCGCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTCTTCCGCG
AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCAGTTAGCAGCGAGCAGACCCGCGCCA
ACAGCCCCACCCGCCCGAGCTGCAGGTGTGGGGCGCGAGAACAAACAGCCTGAGCGAGG
CCGGCGCCGACCGCCAGGGCACCGTGAGCTCAACTCCCCAGATCACCTGTGGCAGC
GCCCTGGTACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTGACACCGCG
CCGACGACACCGTGCTGGAGGAGATGAACCTGCCGGCAAGTGGAAAGCCAAGATGATCG
GCGGGATCGGGGGCTTCATCAAGGTGCGGCACTGACCGAGATCCCCGTGGAGATCTGG
GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCACCCCGTGAACATCATCGGCCGCA
ACCTGCTGACCCAGATCGGCTGCACCTGAACCTCCCCATCAGCCCCATCGAGACGGTGC
CCGTGAAGCTGAAGCCGGGATGGACGCCCAAGGTCAAGCAGTGGCCCTGACCGAGG
AGAAGATCAAGGCCCTGGTGGAGATCTGACCGAGATGGAGAAGGAGGGCAAGATCAGCA
AGATCGGCCCGAGAACCCCTACAACACCCCCGTGTCGCCATCAAGAAGAAGGACAGCA
CCAAGTGGCGCAAGCTGGACTTCCGAGCTGAACAAGCGCACCCAGGACTCTGG
AGGTGCAGCTGGGATCCCCCACCCCGCCGGCTGAAGAAGAAGAAGAGCGTGAACCGTGC
TGGACGTGGCGACGCCCTACTTCAGCGTGCCCCGGACAAGGACTTCCGCAAGTACACCG
CCTTCACCATCCCCAGCATCAACACGAGACCCCGGCATCCGCTACCAGTACAACGTGC
TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATTTCCAGAGCAGCATGACCAAGATCCTGG
AGCCCTTCCGCAAGCAGAACCCGACATCGTATCTACAGTACATGGACGACCTGTACG
TGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCCAGCACC
TGCTGCGCTGGGCTTCACCACCCCGACAAGAACCGACCAAGGAGCCCCCTTCTGT
GGATGGGCTACGAGCTGCACCCGACAAGTGGACCGTGAGCCATCATGCTGCCGAGA
AGGACAGCTGGACCGTGAACGACATCCAGAACAGCTGGTGGCAAGCTGAACCTGGCCAGCC
AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGGGCACCAAGGCC
TGACCGAGGTGATCCCCCTGACCGAGGAGGCCAGCTGGAGCTGCCAGAACCGCGAGA
TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGA
TCCAGAACGGCCAGGGCAGGGCAGTGGACCTACCAAGATCTACAGGAGCCCTTAAGAAC
TGAAGACCGGCAAGTACGCCCGCATGCCGCCACACCAACGACGTGAAGCAGCTGA
CCGAGGCCGTGAGAACAGTGGAGCAGGAGAGCATCGTATCTGGGCAAGATCCCCAAGT
TCAAGCTGCCATCCAGAACGGAGACCTGGAGGCCCTGGTGGAGTACTGGCAGGCCA
CCTGGATCCCCGAGTGGAGTTCTGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAAGC
TGGAGAACGGAGCCCATCGTGGCGCCGAGACCTTACGTGGACGGCGCCCAACCGCG

FIG. 75A

(SEQ ID NO:84)

124/131

AGACCAAGCTGGCAAGGCCGGTACGTGACCGACCGGGCCGGCAGAAGGTGGTGAGCA
TCGCCGACACCACCAACCAGAACGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA
GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGCATCATCCAGGCC
AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG
AGAAGGTGTACCTGGCCTGGGTGCCGCCACAAGGGCATCGCGGCAACGAGCAGGTGG
ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCA
TCGTGATCTACCAAGTACATGGACGACCTGTACGTGGCAGCGCGGCCCTAGGATCGATT
AAAAGCTTCCCAGGGTAGCACCGGTGAATT

FIG. 75B
(SEQ ID NO:84)

125 / 131

Tat_wt_SF162 (wildtype)

ATGGAGCCAGTAGATCCTAGATTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAGA
 CTGCTTGACAAATTGCTATTGAAAAAGTGTGCTTCATTGCCAAGTTGTTCTATAAC
 AAAAGGCTTAGGCATCTCCTATGGCAGGAAGAACGGAGACAGCGACGAAGAGCTCCT
 CCAGACAGTGAGGTTCATCAAGTTCTACCAAAGCAACCCGCTTCCCAGCCCCAAGG
 GGACCCGACAGGCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGA
 TCCAGTCCATTAG

FIG. 76

(SEQ ID NO:85)

Tat_SF162

MEPVDPRLPWKHPGSQPKTACTNCYCKKCCFHCVQCFITKGLGISYGRKKRRQRRAAPPDSE
 VHQVSLPKQPASQPQGDPTGPKESKKVERETEDPVH

FIG. 77

(SEQ ID NO:86)

Tat_SF162_opt

ATGGAGCCCGTGGACCCCCGCTGGAGCCCTGGAAGCACCCGGCAGCCAGCCAAAGAC
 CGCCTGCACCAACTGCTACTGCAAGAAGTGTGCTGCTTCCACTGCCAGGTGTGCTTCATCACC
 AAGGGCCTGGGCATCAGCTACGGCCGCAAGAACGGCCGCCAGCGCCGCCGCC
 CGACAGCGAGGTGCACCAGGTGAGCCTGCCAAGCAGCCGCCAGCCAGCCCCAGGGCG
 ACCCCCACCGGCCCAAGGAGAGCAAGAAGAAGGTGGAGCGCGAGACCGAGACCGACCCCC
 GTGCACTAG

FIG. 78

(SEQ ID NO:87)

Tat_Cys22_SF162_opt

ATGGAGCCCGTGGACCCCCGCTGGAGCCCTGGAAGCACCCGGCAGCCAGCCAAAGAC
 CGCCgGCACCAACTGCTACTGCAAGAAGTGTGCTGCTTCCACTGCCAGGTGTGCTTCATCACC
 AGGGCCTGGGCATCAGCTACGGCCGCAAGAACGGCCGCCAGCGCCGCCGCC
 GACAGCGAGGTGCACCAGGTGAGCCTGCCAAGCAGCCGCCAGCCAGCCCCAGGGCGA
 CCCCACCGGCCCAAGGAGAGCAAGAAGAAGGTGGAGCGCGAGACCGAGACCGACCCCC
 TGCCTAG

FIG. 79

(SEQ ID NO:88)

126 / 131

Alignment GagMod vs GP1_GP2

SUBSTITUTE SHEET (RULE 26)

FIG. 80A

Alignment GagMod vs GP1_GP2

	Section 6				Section 7				Section 8				Section 9			
	Section 6				Section 7				Section 8				Section 9			
	(381)	381	390	400	410	420	430	440	(457)	457	480	490	500	510	520	532
GagMod.	SF2	(381)	CAGCCAGGTGAGCCAGAACTACCCCATCGTGAGAACCTGCAGGGCCAGATGGTCACCAAGGCCATCAGCCCCGC													
GagProtMod.	SF2 (GP1)	(381)	CAGCCAGGTGAGCCAGAACTACCCCATCGTGAGAACCTGCAGGGCCAGATGGTCACCAAGGCCATCAGCCCCGC													
GagProtMod.	SF2 (GP2)	(381)	CAGCCAGGTGAGCCAGAACTACCCCATCGTGAGAACCTGCAGGGCCAGATGGTCACCAAGGCCATCAGCCCCGC													
GagProtMod.	SF2 (GP2)	Consensus	(381)	CAGCCAGGTGAGCCAGAACTACCCCATCGTGAGAACCTGCAGGGCCAGATGGTCACCAAGGCCATCAGCCCCGC												
GagMod.	SF2	(457)	ACCCCTGAACGGCCCTGGGTGAAGGGTGGGGAGGAGAAGGGCCCTTCAGGCCGAGGTGATCCCCATGTTCAGGGCCCTGA													
GagProtMod.	SF2 (GP1)	(457)	ACCCCTGAACGGCCCTGGGTGAAGGGTGGGGAGGAGAAGGGCCCTTCAGGCCGAGGTGATCCCCATGTTCAGGGCCCTGA													
GagProtMod.	SF2 (GP2)	(457)	ACCCCTGAACGGCCCTGGGTGAAGGGTGGGGAGGAGAAGGGCCCTTCAGGCCGAGGTGATCCCCATGTTCAGGGCCCTGA													
GagProtMod.	SF2 (GP2)	Consensus	(457)	ACCCCTGAACGGCCCTGGGTGAAGGGTGGGGAGGAGAAGGGCCCTTCAGGCCGAGGTGATCCCCATGTTCAGGGCCCTGA												
GagMod.	SF2	(533)	533	540	550	560	570	580	590							
GagProtMod.	SF2 (GP1)	(533)	GGGAGGGGCCACCCCCCAGGGACCTGAACACGATGTTGAACACCGTGGGGCCACAGGGCCATGAGATGCT													
GagProtMod.	SF2 (GP2)	(533)	GGGAGGGGCCACCCCCCAGGGACCTGAACACGATGTTGAACACCGTGGGGCCACAGGGCCATGAGATGCT													
GagProtMod.	SF2 (GP2)	Consensus	(533)	GGGAGGGGCCACCCCCCAGGGACCTGAACACGATGTTGAACACCGTGGGGCCACAGGGCCATGAGATGCT												
GagMod.	SF2	(609)	609	620	630	640	650	660	670							
GagProtMod.	SF2 (GP1)	(609)	GAAGGGAGACCATCAACGAGGGAGGCCGAGTGGGACCCCCTGCACCCCCCATGCCGCCGG													
GagProtMod.	SF2 (GP2)	(609)	GAAGGGAGACCATCAACGAGGGAGGCCGAGTGGGACCCCCTGCACCCCCCATGCCGCCGG													
GagProtMod.	SF2 (GP2)	Consensus	(609)	GAAGGGAGACCATCAACGAGGGAGGCCGAGTGGGACCCCCTGCACCCCCCATGCCGCCGG												
GagMod.	SF2	(685)	685	690	700	710	720	730	740							
GagProtMod.	SF2 (GP1)	(685)	CAGATGCGCGAGCCCCGGCACCCCTGCAGGGACATCGGCCGGCACACCCATGCCGCCGG													
GagProtMod.	SF2 (GP2)	(685)	CAGATGCGCGAGCCCCGGCACCCATGCCGCCGGCACACCCATGCCGCCGG													
GagProtMod.	SF2 (GP2)	Consensus	(685)	CAGATGCGCGAGCCCCGGCACACCCATGCCGCCGGCACACCCATGCCGCCGG												

FIG. 80B

128 / 131

Alignment GagMod vs GP1 GP2

	Section 11										Section 12															
GagMod . SF2	(761)	761	770	780	790	800	810	820	830	840	GagMod . SF2	(837)	837	850	860	870	880	890	900	910	912					
GagProtMod . SF2 (GP1)	(761)	ACAACCCCCCCCATCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	CAGCCCCACCCAGCATCTGGACATTCGGCCAGGGGGCCATCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	837	850	860	870	880	GagProtMod . SF2 (GP2)	(761)	ACAACCCCCCCCATTCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	CAGCCCCACCCAGCATCTGGACATTCGGCCAGGGGGCCATCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	837	850	860	870	880	890	900	910	912	
GagProtMod . SF2 (GP2)	(761)	ACAACCCCCCCCATTCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	CAGCCCCACCCAGCATCTGGACATTCGGCCAGGGGGCCATCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	837	850	860	870	880	GagProtMod . SF2 (GP2)	(761)	ACAACCCCCCCCATTCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	CAGCCCCACCCAGCATCTGGACATTCGGCCAGGGGGCCATCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	837	850	860	870	880	890	900	910	912	
Consensus	(761)	ACAACCCCCCCCATTCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	CAGCCCCACCCAGCATCTGGACATTCGGCCAGGGGGCCATCCCCGTGGGAGATCTAACAGGGGTGGATCATCCTGGGCCCTGAAACAGATCGTGGGGATGTA	(837)	837	850	860	870	880	Consensus	(913)	913	920	930	940	950	960	970	980	988	990				
GagMod . SF2	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	913	920	930	940	950	GagMod . SF2	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	913	920	930	940	950	960	970	980	988	990
GagProtMod . SF2 (GP1)	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	913	920	930	940	950	GagProtMod . SF2 (GP2)	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	913	920	930	940	950	960	970	980	988	990
GagProtMod . SF2 (GP2)	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	913	920	930	940	950	GagProtMod . SF2 (GP2)	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(913)	913	920	930	940	950	960	970	980	988	990
Consensus	(913)	ACCCCTGGGGCTGAGCAAGGCCAGGACGTGAAGAACCTGGATGACCCGAGACCTGGCTGTTGGCAGAACGCCAAC	(989)	989	1000	1010	1020	1030	1040	1050	1064	Consensus	(989)	989	1000	1010	1020	1030	1040	1050	1064	1070	1074			
GagMod . SF2	(989)	CCGACTGCAAAGCCATCCITGAAGGCCTCTGGGCCACCCCTGGAGGAGATGATGACCCGAGCTGGCTGCAAGGGCT	(989)	CCGACTGCAAAGCCATCCITGAAGGCCTCTGGGCCACCCCTGGAGGAGATGATGACCCGAGCTGGCTGCAAGGGCT	(989)	989	1000	1010	1020	1030	GagMod . SF2	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	1065	1070	1080	1090	1100	1110	1120	1130	1140	1150
GagProtMod . SF2 (GP1)	(989)	CCGACTGCAAAGCCATCCITGAAGGCCTCTGGGCCACCCCTGGAGGAGATGATGACCCGAGCTGGCTGCAAGGGCT	(989)	CCGACTGCAAAGCCATCCITGAAGGCCTCTGGGCCACCCCTGGAGGAGATGATGACCCGAGCTGGCTGCAAGGGCT	(989)	989	1000	1010	1020	1030	GagProtMod . SF2 (GP2)	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	1065	1070	1080	1090	1100	1110	1120	1130	1140	1150
GagProtMod . SF2 (GP2)	(989)	CCGACTGCAAAGCCATCCITGAAGGCCTCTGGGCCACCCCTGGAGGAGATGATGACCCGAGCTGGCTGCAAGGGCT	(989)	CCGACTGCAAAGCCATCCITGAAGGCCTCTGGGCCACCCCTGGAGGAGATGATGACCCGAGCTGGCTGCAAGGGCT	(989)	989	1000	1010	1020	1030	GagProtMod . SF2 (GP2)	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	1065	1070	1080	1090	1100	1110	1120	1130	1140	1150
Consensus	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	1065	1070	1080	1090	1100	Consensus	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	GGGGGGCCCCGGCCACAAGGGCCGGCTGGCTGGCCAGGGATGAGCCAGGTGACGAACCCGGGACCATCATGATG	(1065)	1065	1070	1080	1090	1100	1110	1120	1130	1140	1150

FIG. 80C

129 / 131

Alignment GagMod vs GP1_GP2

					Section 16	
(1141)	1141	1150	1160	1170	1180	1190
GagMod.	SF2	(1141)	CAGCGGGCAACTTCCGCAACCAGGGAAAGACCGTCAAGTGCTTCAA	CTGGCAAGGAGGGCCACACCGGCCAGGA		1216
GagProtMod.	SF2	(GP1)	(1141)	CAGCGGGCAACTTCCGCAACCAGGGAAAGACCGTCAAGTGCTTCAA	CTGGCAAGGAGGGCCACACCGGCCAGGA	
GagProtMod.	SF2	(GP2)	(1141)	CAGCGGGCAACTTCCGCAACCAGGGAAAGACCGTCAAGTGCTTCAA	CTGGCAAGGAGGGCCACACCGGCCAGGA	
Consensus	(1141)			CAGCGGGCAACTTCCGCAACCAGGGAAAGACCGTCAAGTGCTTCAA	CTGGCAAGGAGGGCCACACCGGCCAGGA	
					Section 17	
(1217)	1217	1230	1240	1250	1260	1270
GagMod.	SF2	(1217)	ACTGCCGCCAAGGGCTGCTGGCCTGCGGCGAAGGGCACCA	ATGAAAGATTGCAC	TGAC	1292
GagProtMod.	SF2	(GP1)	(1217)	ACTGCCGCCAAGGGCTGCTGGCCTGCGGCGAAGGGCACCA	ATGAAAGATTGCAC	TGAGAG
GagProtMod.	SF2	(GP2)	(1217)	ACTGCCGCCAAGGGCTGCTGGCCTGCGGCGAAGGGCACCA	ATGAAAGATTGCAC	TGAGAG
Consensus	(1217)			ACTGCCGCCAAGGGCTGCTGGCCTGCGGCGAAGGGCACCA	ATGAAAGATTGCAC	TGAGAG
					Section 18	
(1293)	1293	1300	1310	1320	1330	1340
GagMod.	SF2	(1293)	CGAGGCCAACCTCCGGCAAGATCTGGCCAGCTACAAGGGC	CCGGCAACTTCTCGAGGCCCCGGAG		1368
GagProtMod.	SF2	(GP1)	(1293)	ACAGGCTTAATTTTAGGAAGATCTGGCCTTCCTACAAGGG	CAAGGGGAAATTTCCTCGAGGAGCCAGAG	
GagProtMod.	SF2	(GP2)	(1293)	ACAGGCTTAATTTTAGGAAGATCTGGCCTTCCTACAAGGG	CAAGGGGAAATTTCCTCGAGGAGCCAGAG	
Consensus	(1293)			ACAGGCTTAATTTTAGGAAGATCTGGCCTTCCTACAAGGG	CAAGGGGAAATTTCCTCGAGGAGCCAGAG	
					Section 19	
(1369)	1369	1380	1390	1400	1410	1420
GagMod.	SF2	(1369)	CCCACGGCCCCCGAGGGAGCCCTCCGGCTTCGGCAAGG	AGGACCCAGCCAGGGAGCCATCG		1444
GagProtMod.	SF2	(GP1)	(1369)	CCAAACAGCCCCACCAAGAGGAGGCTCAGGTGGGGAGGA	ACACTCCCTCTCGAGGAGGGCCGATAG	
GagProtMod.	SF2	(GP2)	(1369)	CCAAACAGCCCCACCAAGAGGAGGCTCAGGTGGGGAGGA	ACACTCCCTCTCGAGGAGGGCCGATAG	
Consensus	(1369)			CCAAACAGCCCCACCAAGAGGAGGCTCAGGTGGGGAGGA	ACACTCCCTCTCGAGGAGGGCCGATAG	
					Section 20	
(1445)	1445	1450	1460	1470	1480	1490
GagMod.	SF2	(1445)	ACAGGGAACTGTACCCCTGACCAAGCAGCCAGGAGCTAA	TTGGCTGGCAAGGAGCCCTCTGTCACAGTAAGGATCGGGGG		
GagProtMod.	SF2	(GP1)	(1445)	ACAGGGAACTGTATCCCTTAACCTCCCTCAAGATCACT	CTTGTCACAGTAAGGATCGGGGG	
GagProtMod.	SF2	(GP2)	(1445)	ACAGGGAACTGTATCCCTCAAGATCACT	CTTGTCACAGTAAGGATCGGGGG	
Consensus	(1445)			ACAGGGAACTGTATCCCTCAAGATCACT	CTTGTCACAGTAAGGATCGGGGG	

FIG. 80D

130 / 131

Alignment GagMod vs GP1 GP2

								Section 21
(1521)	1521	1530	1540	1550	1560	1570	1580	1596
GagMod.	SF2 (1510)							
GagProtMod.	SF2 (GP1) (1521)	CAGCTCAAGGAGGGCGCTCGACACCGGCCGACACCGTGGAGGATGAACCTGCCGGAAAGTGG						
GagProtMod.	SF2 (GP2) (1521)	CAACTCAAGGAAAGGCTCGATACAGGACAGATGATACTAGTTAGAAGAAATGAATTGCCAGGAATATGGA						
Consensus	(1521)	CA CTCAAGGA GGCCTGCTCGA AC GG GC GA AC GT T GA GA ATGAA TGCC GG AA TGGAA						Section 22
(1597)	1597	1610	1620	1630	1640	1650	1660	1672
GagMod.	SF2 (1510)							
GagProtMod.	SF2 (GP1) (1597)	AGCCCCAAGATGATCGGGGGATCAGGTGAGGATCCCCTGGAGATCTGGGG						
GagProtMod.	SF2 (GP2) (1597)	AACCAARAAATGTAAGGGGGATCGGGGATCACCGTGGGGAGTACGACAGATACCTGTAGAAATCTGTGG						
Consensus	(1597)	A CC AA ATGAT GG GGGATCGGGGCTTCATCAAGGT GGCAGTACGACAGAT CC GT GA ATCTG GG						Section 23
(1673)	1673	1680	1690	1700	1710	1720	1730	1748
GagMod.	SF2 (1510)							
GagProtMod.	SF2 (GP1) (1673)	CCACAAAGGCCATCGGCCACCGTGGCTGGGGCCCCACCCCCGTGAAACATCATCGGGCGCAACCTGCTGACCCAGATC						
GagProtMod.	SF2 (GP2) (1673)	ACATAAAAGCTATAGGTACAGTATTAGTAGGACCTAACCTGTCAAACATAATTGGAAAGAAATCTGTGACCCAGATC						
Consensus	(1673)	CA AA GC AT GG AC GT T GT GG CC AC CC GT AACAT AT GG G AA CTG TGACCCAGATC						Section 24
(1749)	1749	1760	1770	1780	1790	1800	1810	1824
GagMod.	SF2 (1510)							
GagProtMod.	SF2 (GP1) (1749)	GGCTGGCACCCATGAACTTCCCCTGAGACGGTGGCCGTGAAGCTGAAGGCCGGATGGACGGCCCCA						
GagProtMod.	SF2 (GP2) (1749)	GGCTGGCACCTTGAACCTTCCCCTGAGACGGTGGCCGTGAAGCTGAAGGCCGGATGGACGGCCCCA						
Consensus	(1749)	GGCTGGCACCCATGAACTTCCCCTGAGACGGTGGCCGTGAAGCTGAAGGCCGGATGGACGGCCCCA						Section 25
(1825)	1825	1830				1847		
GagMod.	SF2 (1510)							
GagProtMod.	SF2 (GP1) (1825)	AGGTCAAGCAAGTGGCCCTGTAA						
GagProtMod.	SF2 (GP2) (1825)	AGGTCAAGCAATGGCAATTGTAA						
Consensus	(1825)	AGGTCAAGCA 1GGCC TGTAA						

FIG. 80E

131 / 131

TataminoSF162.opt

A T G G A G C C C G T G G A C C C C G C C T G G A A G C A C C C C G G A G C C A G C C A
G A C C G C C T T G C A C C A A C T G C T A C T G C A A G G A A G T G C T G C T T C C A C T G C C A G G T G C T T
C A T C A C C A A G G G G C T G G C A T C A G G C T A C G G G G C A A G A A G G G G G C A A G G G G G C

FIG. 81
(SEQ ID NO:89)

Tat_Cys22_SF162

M E P V D P P R L E P W K H P G S Q P K T A G T N C Y C K K C F H C Q V C F I T K G L G I S Y G R K K R R Q R R A P P D S E
V H Q V S L P K Q P A S Q P Q G D P T G P K E S K K V E R E T E T D P V H Z

FIG. 82
(SEQ ID NO:90)

SEQUENCE LISTING

<110> Chiron Corporation

<120> IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
OF VIRUS-LIKE PARTICLES

<130> 1621.100

<140>

<141>

<160> 90

<170> PatentIn Ver. 2.0

<210> 1

<211> 1509

<212> DNA

<213> Human immunodeficiency virus.

<400> 1

atgggtgcga gägcgtcggt attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag gggaaagaa aaatataag taaaaacata tagtatggc aagcaggag 120
ctagaacat tcgcagtcaa tcctggcctg ttagaaacat cagaaggctg cagacaaata 180
ttgggacagc tacagccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
acatgtacaa ccctctattt tgtagatca agatagatc taaaagacac caaggaagct 300
ttagagaaga tagaggaaga gaaaaacaaa agtaagaaaa aggacacagca agcagcagct 360
gcagctggca cagggaaacag cagccaggtc agccaaaatt accctatgt gcagaaccta 420
caggggcaaa tggtacatca ggcataatca cctagaactt taaaatgcatt ggtaaaagta 480
gtagaagaaa aggcttcag cccagaagta ataccatgt tttagcatt atcagaagga 540
gccacccac aagattaaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagatc gcatccatgt 660
catgcagggc ctattgcacc agggcaaatg agagaacca ggggaaatgtc catagcagga 720
actactatgtt cccttcagga acaaatagga tggatgacaa ataattccacc tatcccagta 780
ggagaaatct ataaaagatc gataatcgtt ggattaaata aaatagtaag aatgtatagc 840
cctaccagca ttctggacat aagacaagga ccaaaggaaac cctttagaga ttatgtatagc 900
cggttctata aaactctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
gaaaccttgc tggtccaaaaa tgccaaaccca gattgttataa ctatttttttaa agcattggga 1020
ccagcagcta cactagaaaa aatgtatc gcatgtcagg gagtgcccc accccggccat 1080
aaagcaagag ttttgcttgc agccatgtc agtgcataaa atccatgtt cataatgtatc 1140
cagagaggca atttttaggaa ccaaagaaaag actgttataa gtttcaattt tggcaaaagaa 1200
gggcacatag ccaaaaattt caggccccctt agggaaaagg gctgtggag atgtggaaagg 1260
gaaggacacc aaatgaaaga ttgcactgtc agacaggctt attttttagg gaagatctgg 1320
ccttcctaca agggaaaggcc agggaaattttt cttcagagca gaccagagcc aacagccccca 1380
ccagaagaga gcttcagggtt tggggaggag aaaacaactc ccttcagaca gcaggagcc 1440
atagacaagg aactgtatcc tttaacttcc ctcagatcac tctttggcaa cgaccctcg 1500
tcacaataa 1509

<210> 2

<211> 1845

<212> DNA

<213> Human immunodeficiency virus

<400> 2

atgggtgcga gägcgtcggt attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag gggaaagaa aaatataag taaaaacata tagtatggc aagcaggag 120
ctagaacat tcgcagtcaa tcctggcctg ttagaaacat cagaaggctg cagacaaata 180

ttgggacagc tacagccatc ctttcagaca ggatcagaag aacttagatc attatataat 240
 acatgttcaaa ccctctattt tgtagatcaa agatagatg taaaagacac caaggaagct 300
 tttagagaaga tagaggaaaga gaaaaacaaa agtaagaaaa aggacacagca agcagcagct 360
 gcagctggca caggaaacag cagccaggc agccaaaatt accctatagt gcagaaccta 420
 caggggcaaa tggtagatca gcccataatca cctagaacct taaatgcatt gtaaaagta 480
 gtagaagaaa aggcttcag cccagaagta ataccatgt ttctcagcatt atcagaagga 540
 gccacccac aagattaaa caccatgtca aacacagtgg ggggacatca agcagccatg 600
 caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagagt gcatccagtg 660
 catgcaggc ctattgcacc aggcacaaatg agagaacca ggggaaatgtga catagcagga 720
 actactatgt cccttcagga acaaataatgg tggatgacaa ataatccacc tatcccagta 780
 ggagaaatct ataaaatgt gataatctg ggattaaata aaatagtaag aatgtatagc 840
 ctttaccagea ttctgacat aagacaaggaa cccaaaggaaac ccttttagaga ttatgttagac 900
 cggttctataa aaactctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
 gaaaccttgt tggtcaaaaa tgcaaaaccca gattgtttaa ctattttaaa agcattggga 1020
 ccagcagcta cactagaaga aatgtatgaca gcatgtcagg gagttggggg accccggccat 1080
 aaagcaagag ttttgctga agccatgagc caagtaacaa atccagctaa cataatgtatg 1140
 cagagaggca attttaggaa cccaaagaaag actgtttaagt gtttcaattt tggcaaaagaa 1200
 gggcacatag cccaaaattt cagggccctt agggaaaagg gctgttggag atgttggaaagg 1260
 gaaggacacc aaatgaaaga ttgcactgag agacaggcta attttttagg gaagatctgg 1320
 ctttcctaca agggaaaggcc agggaaatttt cttcagagca gaccagagcc aacagcccc 1380
 ccagaagaga gcttcagggt tggggaggag aaaacaactc ccttcagaa gcaggagccg 1440
 atagacaagg aactgtatcc tttaacttcc ctcagatcac tctttggcaaa cgacccttcg 1500
 tcacaataag gatagggggg caactaaagg aagcttatt agatacagga gcagatgata 1560
 cagtattaga agaaatgtaat ttgccaggaa aatggaaacc aaaaatgtata ggggaaattt 1620
 gaggttttat caaagaaga cagtagatc agatacctgt agaaatctgt ggacataaaag 1680
 ctataggtac agtattatgtt ggacctacac ctgtcaacat aatttggaaaga aatctgttga 1740
 ctcagattgg ttgtacttta aatttcccca tttagtccat taaaactgtt ccagtaaaat 1800
 taaagccagg aatggatggc cccaaagttt agcaatggcc attt 1845

<210> 3
 <211> 4313
 <212> DNA
 <213> Human immunodeficiency virus

<400> 3
 atgggtcgca gagcgtcggt attaagcggg ggagaattt agaaatggg aaaaattcgg 60
 ttaaggccag ggggaaagaa aaaaatataag taaaacata tagttagggc aagcagggag 120
 cttagaacat tcgcagtcaaa tcctggctt ttagaaacat cagaaggctg cagacaata 180
 ttgggacagc tacagccatc ctttcagaca ggatcagaag aacttagatc attatataat 240
 acatgttcaaa ccctctattt tgtagatcaa agatagatg taaaagacac caaggaagct 300
 tttagagaaga tagaggaaaga gaaaaacaaa agtaagaaaa aggacacagca agcagcagct 360
 gcagctggca caggaaacag cagccaggc agccaaaatt accctatagt gcagaaccta 420
 caggggcaaa tggtagatca gcccataatca cctagaacct taaatgcatt gtaaaagta 480
 gtagaagaaa aggcttcag cccagaagta ataccatgt ttctcagcatt atcagaagga 540
 gccacccac aagattaaa caccatgtca aacacagtgg ggggacatca agcagccatg 600
 caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagagt gcatccagtg 660
 catgcaggc ctattgcacc aggcacaaatg agagaacca ggggaaatgtga catagcagga 720
 actactatgt cccttcagga acaaataatgg tggatgacaa ataatccacc tatcccagta 780
 ggagaaatct ataaaatgt gataatctg ggattaaata aaatagtaag aatgtatagc 840
 ctttaccagea ttctgacat aagacaaggaa cccaaaggaaac ccttttagaga ttatgttagac 900
 cggttctataa aaactctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
 gaaaccttgt tggtcaaaaa tgcaaaaccca gattgtttaa ctattttaaa agcattggga 1020
 ccagcagcta cactagaaga aatgtatgaca gcatgtcagg gagttggggg accccggccat 1080
 aaagcaagag ttttgctga agccatgagc caagtaacaa atccagctaa cataatgtatg 1140
 cagagaggca attttaggaa cccaaagaaag actgtttaagt gtttcaattt tggcaaaagaa 1200
 gggcacatag cccaaaattt cagggccctt agggaaaagg gctgttggag atgttggaaagg 1260
 gaaggacacc aaatgaaaga ttgcactgag agacaggcta attttttagg gaagatctgg 1320
 ctttcctaca agggaaaggcc agggaaatttt cttcagagca gaccagagcc aacagcccc 1380
 ccagaagaga gcttcagggt tggggaggag aaaacaactc ccttcagaa gcaggagccg 1440

atagacaagg aactgtatcc tttaacttcc ctcagatcac tctttggcaa cgaccctcg 1500
 tcacaataag gatagggggg caactaaagg aagctctatt agatacagga gcagatgata 1560
 cagtattaga agaaaatgaaat ttgccaggaa aatggaaacc aaaaatgata gggggattg 1620
 gaggttttat caaagtaaga cagtagcgtc agatacctgt agaaatctgt ggacataaag 1680
 ctataggtac agtattagta ggacctacac ctgtcaacat aattgaaaga aatctgttg 1740
 ctcagattgg ttgtacttta aattccccca ttgtcctat tgaaactgtc ccagtaaat 1800
 taaagccagg aatggatggc cccaaagttt agcaatggc attgacagaa gaaaaataa 1860
 aagcattagt agagatgtt acagaaatgg aaaaggaagg gaaaatttca aaaaattggc 1920
 ctgaaaatcc atacaatact ccagtatttgc ctataaaggaa aaaagacagt actaaatgg 1980
 gaaaactagt agatccaga gaacttaata aaagaactca agactctgg gaagttcagt 2040
 taggaatacc acaccccgca gggaaaaaa agaaaaaaatc agtaacagta ttggatgtgg 2100
 gtgtatgcata ctttttagttt ccccttagata aagacttttag aagatatact gcatttacca 2160
 taccttagtat aaacaatgg acacccaggaa tttagatatact gtacaatgtg ctgccacagg 2220
 gatggaaagg atcaccagca atattccaaa gtagcatgac aaaaatctt gaggcatttt 2280
 gaaaacagaa tccagacata gttatctatc aatacatgaa tgatttgtat gtaggatctg 2340
 acttagaaat agggcagcat agaacaaaaaa tagaggaact gagacagcat ctgttgaggt 2400
 ggggattttac cacaccagac aaaaaacatc agaaaagaacc tccatttcctt tgatgggtt 2460
 atgaactcca tcctgataaa tggacagtac agcctataat gtcgcagaa aaagacagct 2520
 ggactgtcata tgacatacag aagttgtgg gaaaattgaa ttggcagttt cagattttatg 2580
 cagggattaa agtaaagcag ttatgtaaac tccttagagg aaccaaagca ctaacagaag 2640
 taataccact aacagaagaa gcagagctag aactggcaga aaacagggag attctaaaag 2700
 aaccagtaca tgaagtatat tatgacccat caaaagactt agtagcagaa atacagaagc 2760
 aggggcaagg ccaatggaca tatcaaattt atcaagagcc attaaaaat ctgaaaacag 2820
 gaaagtatgc aaggatgagg ggtgcccaca ctaatgatgt aaaacagtta acagaggcag 2880
 tgcaaaaagt atccacagaa agcatagtaa tatggggaaa gattctaaa tttaaactac 2940
 ccatacaaaa gggaaacatgg gaagcatggt ggatggagta ttggcaagct acctggattc 3000
 ctgagtgggaa gtttgcataat accccctccct tagtggaaattt atggtaaccat ttagagaaaag 3060
 aaccatagt aggagcagaa actttctatg tagatggggc agctaatagg gagactaaat 3120
 tagggaaaagg aggatatgtt actgacagag caacaaatca gaagactgaa ttcatctgc tttcaggat tccggattag 3240
 aagtaaacat agtaacagac tcacaatatg cattaggaat cattaaagca caaccagata 3300
 agagtgaatc agagtttagtc agtcaataaa tagagcaggat aataaaaaag gaaaaggct 3360
 acctggcatg ggttaccgca cacaaggaa ttggaggaaa tgaacaagta gataaaattag 3420
 tcagtgctgg aatcaggaaa gtactatttt tgaatggaaat agataaggcc caagaagaac 3480
 atgagaaata tcacagtaat tggagagcaa tggcttagta tttaacctg ccacctgttag 3540
 tagaaaaaaa aatagtagcc agctgtgata aatgtcagct aaaaggagaa gccatgcatt 3600
 gacaagttaga ctgtgttca ggaatatggc aactagattg tacacatcta gaaggaaaaa 3660
 ttatccctggt agcagttcat ttagccagtg gatataaga agcagaagttt attccagcag 3720
 agacagggca gggaaacagca tattttctct taaaatttgc aggaagatgg ccagtaaaaa 3780
 caatacatac agacaatggc agcaatttca ccagtactac gtttaaggcc gcctgttggt 3840
 gggcagggat caagcaggaa ttggcatcc cctacaatcc ccaaagtcaa ggagtagtag 3900
 aatctatgaa taatgaatta aagaaaattt taggacaggt aagagatcag gctgaacacc 3960
 ttaagacagc agtacaaatg gcagtattca tccacaattt taaaagaaaa gggggattg 4020
 ggggatacag tgcaggggaa agaataatgtg acataatacg aacagacata caaactaaag 4080
 aactacaaaaa gcaaatttaca aaaatttcaat ttttccgggt ttattacagg gacaacaaag 4140
 atccccctttt gaaaggacca gcaaagctt cttggaaagg tgaagggca gttagtaatac 4200
 aagataataag tgacataaaa gtatgtccaa gaagaaaagc aaaaatcatt agggattatg 4260
 gaaaacagat ggcaggatgtt gattgtgtgg caagtagaca ggatgaggat tag 4313

<210> 4

<211> 1515

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag

<400> 4

gccaccatgg ggcggccgccc cagcgtgctg agcggcgccg agctggacaa gtgggagaag 60
 atccgcctgc gccccggcg caagaagaag tacaagetga agcacatcgt gtgggcccagc 120
 cgcgagctgg agcgcttcgc cgtgaacccc ggcctgtgg agaccagcga gggctgccgc 180
 cagatccctgg gccagctgca gcccagcctg cagaccgca gcgaggagct ggcgcagcctg 240
 tacaacacccg tggccaccct gtactgcgtg caccagcga tcgacgtcaa ggacacccaag 300
 gagggccctgg agaagatcga ggaggaggcag aacaagtcaa agaagaaggc ccagcaggcc 360
 gcccggcccg cccgcacccgg caacacgcg caggtgagcc agaactaccc catcgtgcag 420
 aacctgcagg gccagatggt gcacccggcc atcggccccc gcacccctgaa cgcctgggtg 480
 aagggttgtgg aggagaaggc cttcagcccc gagggtatcc ccatgttcaag cgcctggagg 540
 gaggggcgcca ccccccaggta cctgaacacg atgttgaaca cctgtggcg ccaccaggcc 600
 gccatgcaga tgctgaagga gaccatcaac gaggaggccg cccgatggaa cccgcgtgcac 660
 cccgtgcacg cccggccccc gccccccggc cagatgcgcg agcccccggg cagcgacatc 720
 gcccggccacca ccagcaccct gcaggaggcag atcggcttggaa tgaccaacaa ccccccattc 780
 cccgtggcg agatctacaa ggggtggatc atccctggcc tgaacaagat cgtgcggatg 840
 tacagccccca ccagcatctt ggacatccgc caggggccca aggagccctt cccgcactac 900
 gtggaccegt tctacaagac cctgcgcgtt gaggaggccca gccaggacgt gaagaactgg 960
 atgacccgaga ccctgttggt gcagaacgcgc aaccccgact gcaagaccat cctgaaggct 1020
 ctccggccccc cggccaccctt ggaggaggatg atgaccgcctt gccaggccgt gggcgcccc 1080
 gcccacaagg cccgcgtgtt gggcgaggccg atgagccagg tgaccaaccc ggcgaccatc 1140
 atgatgcagc gggcaacctt ccgcaaccag cggaagaccg tcaagtgcctt caactgcggc 1200
 aaggaggggcc acaccggccag gaactgcgcg gccccccgca agaagggttg ctggcgctgc 1260
 gggcgcgagg gccaccagat gaaggactgc accgagccgc aggccaactt cctggcgaaag 1320
 atctggccca gctacaaggg cccggccccc gacttgcgtc agagccccc cgagcccacc 1380
 gcccccccccc aggagagctt ccgttgcggc gaggagaaga ccaccccccagg ccagaaggcag 1440
 gagcccatcg acaaggagct gtacccctg accagctgc gcagctgtt cggcaacgac 1500
 cccagcagcc agtaa 1515

<210> 5
 <211> 1853
 <212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag-protease

<400> 5

gccaccatgg ggcggccgccc cagcgtgctg agcggcgccg agctggacaa gtgggagaag 60
 atccgcctgc gccccggcg caagaagaag tacaagetga agcacatcgt gtgggcccagc 120
 cgcgagctgg agcgcttcgc cgtgaacccc ggcctgtgg agaccagcga gggctgccgc 180
 cagatccctgg gccagctgca gcccagcctg cagaccgca gcgaggagct ggcgcagcctg 240
 tacaacacccg tggccaccct gtactgcgtg caccagcga tcgacgtcaa ggacacccaag 300
 gagggccctgg agaagatcga ggaggaggcag aacaagtcaa agaagaaggc ccagcaggcc 360
 gcccggcccg cccgcacccgg caacacgcg caggtgagcc agaactaccc catcgtgcag 420
 aacctgcagg gccagatggt gcacccggcc atcggccccc gcacccctgaa cgcctgggtg 480
 aagggttgtgg aggagaaggc cttcagcccc gagggtatcc ccatgttcaag cgcctggagg 540
 gaggggcgcca ccccccaggta cctgaacacg atgttgaaca cctgtggcg ccaccaggcc 600
 gccatgcaga tgctgaagga gaccatcaac gaggaggccg cccgatggaa cccgcgtgcac 660
 cccgtgcacg cccggccccc gccccccggc cagatgcgcg agcccccggg cagcgacatc 720
 gcccggccacca ccagcaccctt gcaggaggccg atcggcttggaa tgaccaacaa ccccccattc 780
 cccgtggcg agatctacaa ggggtggatc atccctggcc tgaacaagat cgtgcggatg 840
 tacagccccca ccagcatctt ggacatccgc caggggccca aggagccctt cccgcactac 900
 gtggaccegt tctacaagac cctgcgcgtt gaggaggccca gccaggacgt gaagaactgg 960
 atgacccgaga ccctgttggt gcagaacgcgc aaccccgact gcaagaccat cctgaaggct 1020
 ctccggccccc cggccaccctt ggaggaggatg atgaccgcctt gccaggccgt gggcgcccc 1080
 gcccacaagg cccgcgtgtt gggcgaggccg atgagccagg tgaccaaccc ggcgaccatc 1140
 atgatgcagc gggcaacctt ccgcaaccag cggaagaccg tcaagtgcctt caactgcggc 1200
 aaggaggggcc acaccggccag gaactgcgcg gccccccgca agaagggttg ctggcgctgc 1260
 gggcgcgaaag gacaccaaattt gaaagattgc actgagagac aggctaattt tttagggaaag 1320

atctggcctt cctacaaggg aaggccaggg aattttcttc agagcagacc agagccaaca 1380
 gccccaccag aagagagctt caggtttggg gaggagaaaa caactccctc tcagaagcag 1440
 gagccgatac acaaggaact gtatccctta acttccctca gatcactt tggcaacgac 1500
 ccctcgac acgtaaaggatc ggcggccagc tcaaggagac gctgtcgac accggcggcg 1560
 acgacaccgt gctggggag atgaacacctc cgggcaagtg gaagcccaag atgatcggcg 1620
 ggatcggggg cttcatcaag gtgcggcagt acgaccat cccctggag atctgcggcc 1680
 acaaggccat cggcacccgt ctggtgggcc ccacccctgt aaacatcatc gggcgaacc 1740
 tgctgacccca gatcggtcgc accctgaact tccccatcag ccccatcag acgggtggcc 1800
 tgaagctgaa gcccgggatc gacggcccca aggtcaagca gtggccctg taa 1853

<210> 6

<211> 4319

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
 HIV-Gag-polymerase

<400> 6

gccaccatgg gcgccgcgc cagcgtctg agcggcggcg agctggacaa gtgggagaag 60
 atccgcctgc gccccgggg caagaagaag tacaagctga agcacatcg tggggccagc 120
 cgcgagctgg agcgttcgc cgtgaacccc ggcctgtgg agaccagcga gggctgccc 180
 cagatcctgg gccagctgca gcccagectg cagacccgca gcgaggagct gccgcgtctg 240
 tacaacaccc tggccacccct gtactcgctg caccagcgtc tgcacgtcaa ggacaccaag 300
 gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
 gcccggcccg cggcacccgg caacagcage caggtgagcc agaactaccc catcggtcag 420
 aacctgcagg gccagatgtt gcacccaggcc atcagccccc gcacccctgaa cggctgggtg 480
 aagggtgtgg aggagaaggc cttcagcccc gaggtgatcc ccatgttcag cggccctgagc 540
 gagggcgcca ccccccagga cctgaacacg atgttgaaca cctgtggcgcc ccaccaggcc 600
 gccatgcaga tgcgtgaagga gaccatcaac gaggaggccg ccgagtggga cccgcgtcac 660
 cccgtgcacg cggccccat cggcccccgc cagatcgccg agccccggcg cagcgacate 720
 gccccgacca ccagcacccct gcaggagcag atcggctgga tgaccaacaa ccccccacatc 780
 cccgtggcg agatctacaa gcggtggatc atcctggggc tgaacaaat cgtgcggatg 840
 tacagcccca ccagcatect ggacatccgc cagggcccca aggagccctt cccgcactac 900
 gtggaccgct tctacaagac cctgcgcgtc gagcaggccca gccaggacgt gaagaacttg 960
 atgaccgaga ccctgttgtt gcagaacgc aaccccgact gcaagaccat cctgaaggct 1020
 ctgcggccccc cggccacccct ggaggagatc atgaccgcgt gccaggcggt gggggcccc 1080
 ggccacaagg cccgcgtgtc ggccgaggcg atgagccagg tgacgaaccc ggcgaccatc 1140
 atgatgcagc gcccacactt ccccaaccag cggaaagaccc tcaagtgcgtt caactgcggc 1200
 aaggaggggcc acacccgcag gaactgcgcg ccccccgcga agaaggcgct ctggcgctgc 1260
 gggccgcaag gacacccaaat gaaagattgc actgagagac aggttaattt tttagggaaag 1320
 atctggcctt cctacaaggg aaggccaggg aattttcttc agagcagacc agagccaaca 1380
 gccccaccag aagagagctt caggtttggg gaggagaaaa caactccctc tcagaagcag 1440
 gagccgatac acaaggaact gtatccctta acttccctca gatcactt tggcaacgac 1500
 ccctcgac acgtaaaggatc ggcggccagc tcaaggagac gctgtcgac accggcggcg 1560
 acgacaccgt gctggggag atgaacacctc cgggcaagtg gaagcccaag atgatcggcg 1620
 ggatcggggg cttcatcaag gtgcggcagt acgaccat cccctggag atctgcggcc 1680
 acaaggccat cggcacccgt ctggtgggcc ccacccctgt aaacatcatc gggcgaacc 1740
 tgctgacccca gatcggtcgc accctgaact tccccatcag ccccatcag acgggtggcc 1800
 tgaagctgaa gcccgggatc gacggcccca aggtcaagca gtggccctg accgaggaga 1860
 agatcaaggg cctgggtggag atctgcaccc agatggagaa ggagggcaag attagcaaga 1920
 tcggcccccga gaacccctac aacacccccc ttttcgcctt caagaagaag gacagcacca 1980
 agtggcgcaa gctggtgac ttccgcgagc tgaacaaagcg caccaggac ttctgggagg 2040
 tgcagctggg catccccac cccgcggccc tgaagaagaa gaagagcgtg accgtgcgtt 2100
 acgtggcgca cgcctacttc agcgtgcccc tggacaagga cttccgcac tacaccgcct 2160
 tcacccatccc cagcatcaac aacgagaccc cccgcacccgt ctaccaggatc aacgtgcgtc 2220
 cccaggccgt gaagggcagc cccgcacatc tccagagcag catgaccaag atcctggagc 2280
 cttccgcacaa gcacccatc gacatcgta tttaccaggatc catggacgac ctgtacgtgg 2340

gcagcgacct ggagatcggc cagcacccgca ccaagatcga ggagctgcgc cagcacctgc 2400
 tgcgctgggg cttcaccacc cccgacaaga agcaccagaa ggagccccccc ttccctgtgga 2460
 tgggctacga gctgcacccc gacaagtggc ccgtgcagcc catcatgtcg cccgagaagg 2520
 acagctggac cgtgaacgcg atccagaagc tggtgggca gctgaactgg gccagccaga 2580
 tctacgcccgg catcaaggtg aagcagctgt gcaagctgt ggcggccacc aaggccctga 2640
 ccgaggttat cccccctgacc gaggaggccg agctggagct ggccgagaac cgcgagatcc 2700
 tgaaggaggcc cgtgcacccg gtgtactacg accccagca ggacctggg gcccggatcc 2760
 agaaggcaggc ccaggggccag tggacctacc agatctacca ggagcccttc aagaacctga 2820
 agaaggcggca gtacccccgc atgcggccggc cccacacca cgacgtgaaag cagctgaccg 2880
 aggccgtgca gaaggtgagc accgagagca tctgtatcg gggcaagatc cccaagttca 2940
 agetccccat ccagaaggag acctggggg cctggtgat ggagtaactgg cagggccacct 3000
 ggatccccca gtgggagttc gtggaaacacc cccccctgtt ggagctgtgg taccagctgg 3060
 agaaggagcc categtgggc gcccggaccc tctacgtgga cggccggcc aaccgcgaga 3120
 ccaagctggg caaggccggc taegtgaccg accggccggc ccagaagggtg gtgagcatcg 3180
 ccgacaccac caaccagaag accgagctgc aggccatcca cctggccctg caggacagcg 3240
 gcctggaggt gaacatcgtg accgacagcc agtacgcct ggccatcatc caggccccagc 3300
 ccgacaagag cgagagcgag ctggtgagcc agatcatega gcagctgatec aagaaggaga 3360
 aggtgtactt ggcctgggtt cccggccaca agggcategg cggcaacgcg caggtggaca 3420
 agctggtgag cggccggcatt cggcaagggtc tggttctgaa cggcategac aaggcccagg 3480
 aggagcacga gaagtaccac agcaactggc ggcgcattgc cagcacttc aacctgcccc 3540
 ccgtgggtgc caaggagatc gtggccagct ggcacaagtgc cagctgaag ggcgaggcca 3600
 tgcacggccca ggtggactgc agccccggca tctggcaactt ggactgcacc cacctggagg 3660
 gcaagatcat cctggtgccca gtgcacgtgg ccagcggtt catcgaggcc gaggtgatcc 3720
 ccgcccggagac cgcccgaggag accgcctactt tctgtctgaa gctggccggc cgctggcccg 3780
 tgaagaccat ccacaccgac aacggcagca acttcaccag caccaccgtg aaggccgcct 3840
 gctgggtggc cggcatcaag caggagttcg gcatccccca caaccccccag agccaggggcg 3900
 tggtgagag catgaacaac gagctgaaga agatcategg ccaggtgcgc gaccaggccg 3960
 agcacctgaa gaccggcgtg cagatggccg tggttcatcca caacttcaag cgcaaggggcg 4020
 gcatcgccgg ctacagcgcc ggcgagcgca tctgtggacat ctcgcacc gacatccaga 4080
 ccaaggagct gcagaaggcag atcacaaga tccagaactt ccgcgtgtac taccgcgaca 4140
 acaaaggaccc cctgtggaaagg ggccccggca agtgcgtgtg gaagggcgag ggcgcccgtgg 4200
 tcatcccgagaa caacagcgac atcaagggtgg tgcccgccca caaggccaaatc atcatcccg 4260
 actacggccaa gcagatggcc ggccgacgact gctggcccg cgcgcaggac gaggactag 4319

<210> 7

<211> 2031

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag/HCV-core fusion polypeptide

<400> 7

gcccaccatgg ggcggccggc cagcgtgctg agcggccggcg agctggacaa gtgggagaag 60
 atccgcctgc gccccggcg caagaagaag tacaagctga agcacategt gtggggccagc 120
 cgcgagctgg agcgcttcgc cgtgaacccc ggcctgtgg agaccagcgaa gggctggccgc 180
 cagatcttgg gccagctgc gcccagccctg cagacccggca ggcggaggct ggcgcagccctg 240
 tacaacaccgg tggccaccctt gtactgcgtg caccagcgca tcgcacgtcaa ggacaccaag 300
 gaggccctgg agaagatcga ggaggaggcag aacaagtcca agaagaaggc ccagcaggcc 360
 gcccggcccg ccggcaccgg caacagcgac caggtgagcc agaactaccc catcggtcag 420
 aacctgcagg gccagatggt gcaccaggcc atcagcccccc gcaccctgtt ggcctgggtt 480
 aagggtgggg aggagaaggc cttcagcccccc gaggtgatcc ccatgtttag cgcctggggc 540
 gaggggccca cccccccaggca cctgaacacgg atgttgaaca ccgtggccgg ccaccaggcc 600
 gccatgcaga tgctgaagga gaccatcaac gaggaggccg ccgagtggga ccgcgtgcac 660
 cccgtgcacg cggcccccat cggcccccggc cagatgcgcg agccccggc cagcgacatc 720
 gccggccacca ccagcaccctt gcaggaggcag atcggctggaa tgaccaacaa cccccccatc 780
 cccgtggggcg agatctacaa gcggtggatc atcctggccca tgaacaagat cgtgcggatg 840
 tacagccccca ccagcatccgc ggacatccgc cagggccctt cgcgcactac 900

gtggacgcgt tctacaagac cctgcgcgct gagcaggcca gccaggacgt gaagaactgg 960
 atgaccgaga ccctgtggc gcagaacgc aaccccgact gcaagaccat cctgaaggct 1020
 ctggcccccg cggccacccct ggaggagatg atgaccgcgt gccaggcgt gggcgcccc 1080
 ggcccacaagg cccgegtgct ggccgaggcg atgagccagg tgacgaaacc cgcgaccate 1140
 atgatgcagc gggcaactt cgcaccaacc cggaaagaccc tcaagtgcctt caactgcggc 1200
 aaggaggggcc acacccgcg gaactgcgcg gccccccgca agaagggctg ctggcgctgc 1260
 ggccgcgagg gccaccagat gaaggactgc accgagcgcc aggcaacatt cctgggcaag 1320
 atctggccca getacaagggg cccggccggc aacttcctgc agagccggcc cgagccacc 1380
 gcccccccg aggagagctt ccgttcggc gaggagaaga ccaccccgag ccagaaggcag 1440
 gagccccatcg acaaggagct gtacccctg accagcctgc gcagcctgtt cggcaacgcac 1500
 cccagcagcc agtgcacgaa tcctaaacact caaagaaaaa acaaacgtaa caccacccgt 1560
 cgccccacagg acgtcaagtt cccgggtggc ggtcagatcg ttggtgaggt ttacttgg 1620
 cccgcgcagg gcccttagatt gggtgtgcgc ggcacgagaa agactccga ggggtcgaa 1680
 cctcgaggta gacgtcagcc tatccccaaag gtcgtcgcc cggaggcgag gacctgggct 1740
 cagccccgggt acccttggcc cctctatggc aatgagggtt ggggtgggc gggatggctc 1800
 ctgtetcccc gtggctctcg gcctagctgg gggccccacag acccccccggc taggtcgcc 1860
 aatttgggta aggtcatcga tacccttacg tgccggcttcg ccgacctcat ggggtacata 1920
 ccgcctegtcg gcccctctt tggaggcgct gccaggccc tggcgcatgg cgtccgggtt 1980
 ctggaagacg gcgtgaacta tgcaacacagg aaccttcctg gttgtctta g 2031

<210> 8

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag/HCV-Core fusion polypeptide

<400> 8

atgggtgcga gagcgtcggt attaagcggg ggagaattag ataaatggga aaaaattcgg 60
 ttaaggccag ggggaaaagaa aaaatataag taaaacata tagtatggc aagcaggagg 120
 ctagaacat tcgcagtcaa tcctggcctg ttagaaacat cagaaggctg cagacaaata 180
 ttgggacagc tacagccatc cttcagaca ggatcagaag aacttagatc attatataat 240
 acagtagcaa cccttattt tgtacatcaa aggatagatg taaaagacac caaggaagct 300
 ttagagaaga tagaggaaga gaaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
 gcagctggca caggaaacag cagccaggc agccaaaatt accctatagt gcagaaccta 420
 caggggcaaa tggtacatca ggcataatca cctagaacct taaatgcatt ggtaaaagta 480
 gtagaagaaa aggctttcag cccagaagta ataccatgt tttcagcatt atcagaagga 540
 gcccacccac aagattttaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
 caaatgttaa aagagactat caatgaggaa gtcgcagaat gggatagatg gcatccagt 660
 catgcaggcc ctattgcacc aggcacaaatg agagaaccaa ggggaagtga catagcagga 720
 actacttagt cccttcagga acaaataggg tggatgacaa ataattccacc tatcccagta 780
 ggagaaaatctt ataaaaagatg gataatccctg ggataaaat aaatagtaag aatgtatagc 840
 cttaccagca ttctggacat tagacagga ccaaaggaaac ccttttagaga ttatgtac 900
 cggttctata aaactctaag agccgaacaa gtttcacagg atgtaaaaaa ttggatgaca 960
 gaaaccttgc tggtccaaaaa tgccaaaccca gattgtaaa ctattttaa acattggga 1020
 ccagcagcta cactagaaga aatgtgaca gcatgtcagg gatgtggggg acccgccat 1080
 aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgtg 1140
 cagagaggca attttagaa ccaaagaaatg actgttaagt gttcaattt tgccaaagaa 1200
 gggcacatag cccaaaaattt cagggccccctt agggaaaagg gctgtggag atgtggagg 1260
 gaaggacacc aaatgaaaaga ttgcactgag agacaggctt attttttagg gaagatctgg 1320
 cttccctaca agggaaaggcc agggaaatttt cttcagagca gaccagagcc aacagcccc 1380
 ccagaagaga gtttcagggtt tggggaggag aaaacaactc ccttcagaa gcaggagccg 1440
 atagacaagg aactgtatcc tttaacttcc ctcagatcac tctttggcaa cgacccctcg 1500
 tcacagtcgaa cgaatccaa acctcaaaga aaaaacaac gtaacaccaa ccgtcgccca 1560
 caggacgtca agttcccccggg tggcggtcag atcggtggg gatgttactt gtgcgcgc 1620
 agggccctta gattgggtgt gcgcgcgacg agaaagactt ccgagcggc gcaacctcg 1680
 ggttagacgtc agcctatccc caaggctcgt cggcccgagg gcaggacctg ggctcagccc 1740

```

gggttacccctt ggcggccatca tggcaatggag ggctgggggt gggcgggatg gtcctgtct 1800
ccccgtggct ctcggcctag ctggggggccc acagaccccc ggcttaggtc ggcgaatttg 1860
ggtaaggctca tcgataccct tacgtgcggc ttgcggcace tcatggggta cataccgctc 1920
gtcgccggcc ctcttgagg cgctgccagg gcccctggcgc atggcgtccg ggttctggaa 1980
gacggcgtga actatgcaac agggAACCTT cctgggttgc ttttag 2025

```

<210> 9
<211> 1268
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic Gag
common region

<210> 10
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: HIV-Gag
peptide p7G

<400> 10
Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu
1 5 10 15

Glu Ala Ala Glu
20

<210> 11
<211> 30
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer GAG5

<400> 11

aagaattcca tgggtgcgag agcgtcggtta

30

<210> 12

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer
pS5-SAL3

<400> 12

attcgtcgac tgtgacgagg ggtcgttgcc

30

<210> 13

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer
CORESAL5

<400> 13

atttgcgac gaatcctaaa cctcaaagaa aaac

34

<210> 14

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer 173CORE

<400> 14

tattggatcc taagagcaac caggaaggtt c

31

<210> 15

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer MS65

<400> 15

cgaccatcat ggatgcagcg c

21

<210> 16

<211> 30

<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer MS66

<400> 16
aggattcgtc gagtcgctgc tggggtcgtt

30

<210> 17
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer XPANXNF

<400> 17
gcacgtggc cccggcgctc tagagc

26

<210> 18
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer XPANXNR

<400> 18
gtcttagagg cgccggggccc acgtgc

26

<210> 19
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: HIV p55 Gag
Major Homology Region

<400> 19
Asp Ile Arg Gln Gly Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg
1 5 10 15

Phe Tyr Lys Thr
20

<210> 20
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic p55
Gag Major Homology Region

<400> 20
gacatccgcc agggccccaa ggagcccttc cgcgactacg tggaccgctt ctacaagacc 60

<210> 21
<211> 15

<212> PRT

<213> Human immunodeficiency virus

<400> 21

Ala	Pro	Thr	Lys	Ala	Lys	Arg	Arg	Val	Val	Gln	Glu	Lys	Arg
1.					5			10				15	

<210> 22

<211> 5

<212> PRT

<213> Human immunodeficiency virus

<400> 22

Lys	Ala	Lys	Arg	Arg
1		5		

<210> 23

<211> 4

<212> PRT

<213> Human immunodeficiency virus

<400> 23

Arg	Glu	Lys	Arg
1			

<210> 24

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of
mut7.SF162 cleavage site

<400> 24

Ala	Pro	Thr	Lys	Ala	Ile	Ser	Ser	Val	Val	Gln	Ser	Glu	Lys	Ser
1					5			10				15		

<210> 25

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of
mut8.SF162 cleavage site

<400> 25

Ala	Pro	Thr	Ile	Ala	Ile	Ser	Ser	Val	Val	Gln	Ser	Glu	Lys	Ser
1					5			10				15		

<210> 26

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of
mut.SF162 cleavage site

<400> 26

Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Ser
1 5 10 15

<210> 27

<211> 15

<212> PRT

<213> Human immunodeficiency virus

<220>

<223> Description of Artificial Sequence: aa of native
cleavage site in US4

<400> 27

Ala Pro Thr Gln Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
1 5 10 15

<210> 28

<211> 5

<212> PRT

<213> Human immunodeficiency virus

<220>

<223> Description of Artificial Sequence: aa of second
cleavage site in US4

<400> 28

Gln Ala Lys Arg Arg
1 5

<210> 29

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of mut.US4
cleavage site

<400> 29

Ala Pro Thr Gln Ala Lys Arg Arg Val Val Gln Arg Glu Lys Ser
1 5 10 15

<210> 30

<211> 1419

<212> DNA

<213> Human immunodeficiency virus

<400> 30

gtagaaaaat tgggttcac agtctattat ggggtacctg tggaaaga agcaaccacc 60
 actctatttt gtgcacccaga tgctaaagcc tatgacacag aggtacataa tgtctggcc 120
 acacatgcct gtgtacccac agaccctaac ccacaagaaa tagtattgaa aatgtgaca 180
 gaaaatttta acatgtggaa aaataacatg tggaaacaga tgcacggga tataatcagt 240
 ttatggatc aaagtctaa gccatgtgt aagttaccc cactctgtgt tactctacat 300
 tgcactaatt tgaagaatgc tactaatacc aagagttagta attggaaaga gatggacaga 360
 ggagaaataa aaaattgttc tttcaagggtc accacaagca taagaataa gatgcagaaa 420
 gaatatgcac tttttataa acttgatgtt gttaccaatg ataatgataa tacaagctat 480
 aaattgataa attgtacac ctcagtcatt acacaggctt gtccaaaggtt atccctttgaa 540
 ccaattcccc tacattattt tgccccggct ggttttgcga ttctaaagtgt taatgataag 600
 aagttcaatg gatcagggacc atgtacaaat gtcagcacag tacaatgtac acatggaaatt 660
 aggccagtag tgcactaattcattt atggcgatgtt aatggcgatgtc tagcagaaga aggggttagta 720
 attagatctg aaaatttcac agacaatgtt aaaaactataa tagtacagctt gaaggaaatct 780
 gtagaaatttta attgtacaag acctaacaat aatacaagaa aaagtataac tataggaccg 840
 gggagagcat ttatgcaac aggagacata ataggagata taagacaagc acattgtaac 900
 attagtggag aaaaatggaa taacacttta aaacagatgtt ttacaaaattt acaagcaca 960
 ttggaaataa aaacaatagt cttaagcaa tcctcaggag gggaccaga aattgtatg 1020
 cacagtttta attgtggagg ggaatttttc tactgttattt caacacagctt tttaatagt 1080
 acttggaaata atactatagg gccaataaac actaatggaa ctatcacact cccatgcaga 1140
 ataaaacaaa ttataaacag gtggcaggaa gtagggaaaag caatgtatgc ccctccccatc 1200
 agaggacaaa tttagatgttc atcaaataattt acaggactgtc tattaacaag agatgggttgt 1260
 aaagagatca gtaacaccac cgagatcttc agacctggag gtggagatataatg gaggacaaat 1320
 tggagaagtg aatttatataa atataaagta gtaaaaattt agccattttt agtagcacc 1380
 accaaggca agagaagagt ggtgcagaga gaaaaaagag cagtgcacgtt agagactatg 1440
 1419

<210> 31

<211> 1932

<212> DNA

<213> Human immunodeficiency virus

<400> 31

gtagaaaaat tgggttcac agtctattat ggggtacctg tggaaaga agcaaccacc 60
 actctatttt gtgcacccaga tgctaaagcc tatgacacag aggtacataa tgtctggcc 120
 acacatgcct gtgtacccac agaccctaac ccacaagaaa tagtattgaa aatgtgaca 180
 gaaaatttta acatgtggaa aaataacatg tggaaacaga tgcacggga tataatcagt 240
 ttatggatc aaagtctaa gccatgtgt aagttaccc cactctgtgt tactctacat 300
 tgcactaatt tgaagaatgc tactaatacc aagagttagta attggaaaga gatggacaga 360
 ggagaaataa aaaattgttc tttcaagggtc accacaagca taagaataa gatgcagaaa 420
 gaatatgcac tttttataa acttgatgtt gttaccaatg ataatgataa tacaagctat 480
 aaattgataa attgtacac ctcagtcatt acacaggctt gtccaaaggtt atccctttgaa 540
 ccaattcccc tacattattt tgccccggct ggttttgcga ttctaaagtgt taatgataag 600
 aagttcaatg gatcagggacc atgtacaaat gtcagcacag tacaatgtac acatggaaatt 660
 aggccagtag tgcactaattcattt atggcgatgtt aatggcgatgtc tagcagaaga aggggttagta 720
 attagatctg aaaatttcac agacaatgtt aaaaactataa tagtacagctt gaaggaaatct 780
 gtagaaatttta attgtacaag acctaacaat aatacaagaa aaagtataac tataggaccg 840
 gggagagcat ttatgcaac aggagacata ataggagata taagacaagc acattgtaac 900
 attagtggag aaaaatggaa taacacttta aaacagatgtt ttacaaaattt acaagcaca 960
 ttggaaataa aaacaatagt cttaagcaa tcctcaggag gggaccaga aattgtatg 1020
 cacagtttta attgtggagg ggaatttttc tactgttattt caacacagctt tttaatagt 1080
 acttggaaata atactatagg gccaataaac actaatggaa ctatcacact cccatgcaga 1140
 ataaaacaaa ttataaacag gtggcaggaa gtagggaaaag caatgtatgc ccctccccatc 1200
 agaggacaaa tttagatgttc atcaaataattt acaggactgtc tattaacaag agatgggttgt 1260
 aaagagatca gtaacaccac cgagatcttc agacctggag gtggagatataatg gaggacaaat 1320
 tggagaagtg aatttatataa atataaagta gtaaaaattt agccattttt agtagcacc 1380
 accaaggca agagaagagt ggtgcagaga gaaaaaagag cagtgcacgtt agagactatg 1440
 ttccttgggt tcttgggagc agcaggaaagc actatggggc cacggctact gacgctgacg 1500
 gtacaggccca gacaattttt gtctggatata gtcacacagc agaacaattt gctgagatct 1560
 attggggcgc aacagcatctt gttcaactc acagtctggg gcatcaagca gtcggcaggca 1620
 agagtccctgg ctgtggaaag atacctaaag gatcaacagc tccttagggat ttgggggttgc 1680

tctggaaaac	tcatttgcac	cactgctgtg	ccttggaaatg	ctagttggag	taataaaatct	1740
ctggatcaga	tttgaataa	catgacacctg	atggagtg	ggagagaaaat	tgacaattac	1800
acaaacttaa	tatacacctt	aattgaagaaa	tcgcagaacc	aacaagaaaa	aatgaacaa	1860
gaattattag	aattggataa	gtgggcaagt	tttgtggatt	ggtttgcacat	atcaaaaatgg	1920
ctgtggata	ta					1932

<210> 32
<211> 2457
<212> DNA
<213> Human immunodeficiency virus

<400> 32
gtaaaaaaaaat tgtgggtcac agtctattat ggggtacctg tggaaaaga agcaaccacc 60
actctatttt gtgcatacaga tgctaaagcc t'atgacacag aggtacataa tgcacccccc 120
acacatgcct gtgtacccac agaccctaac ccacaagaaa tagtatttggaa aatgtgaca 180
aaaaatttta acatgtggaa aaataaacatg gtagaacaga tgcatacgaggaa tataatcagt 240
ttatgggatc aaagtctaaa gccatgtgt aagttAACCC cactctgtgt tactctacat 300
tgcactaatt tgaagaatgc tactaatacc aagagttagta attggaaaga gatggacaga 360
ggagaaaataa aaaattgttc ttcaagggtt accacaagca taagaaataa gatgcagaaaa 420
aatatgcac ttttttataa acttgatgtt gtagccaatag ataatgataa tacaagctat 480
aaattgataa attgtacac ctcagtcatt acacagggct gtcacaaaggt atcccttgaa 540
ccaattccca tacattattt tgccccggct gtttttgcga ttctaaagtg taatgataaag 600
aagtcaatg gtcaggacc atgtacaaat gtcagcacag tacaatgtac acatggaaatt 660
aggccagtag tgcactca attgtgtt aatggcagtc tagcagaaga aggggttaga 720
attagatctc aaaattcac agacaatgtc aaaactataa tagtacagct gaaggatct 780
gtagaaaatta atttacaag acctaacaat aatacaagaa aaagtataac tataggaccg 840
gggagagcat ttatgcaac aggagacata ataggagata taagacaagc acatgttaac 900
attagtggag aaaaatggaa taacactttt aaacagatag ttacaaaatt acaagcacaa 960
tttgggataa aaacaatagt cttaagcaa tcctcaggag gggacccaga aattgtatag 1020
cacagttta attgtggagg ggaatttttc tactgttaatt caacacagct ttttaatagt 1080
acttggataa atactatagg gccaataaac actaatggaa ctatcacact cccatgcaga 1140
ataaaaacaaa ttataaacag gtggcaggaa gtagaaaaag caatgtatgc ccctcccatc 1200
agaggacaaa ttagatgttc atcaaataattt acaggactgc tattaacaag agatgggtgt 1260
aaagagatca gtaacaccac cgagatctt agacctggag gtggagatata gagggacaat 1320
tggagaagtg aattatataa atataaaagta gtaaaaaattt agcatttagg agtagcaccc 1380
accaaggc aaagagaagat ggtgcagaga gaaaaaaagag cagtgcgcg aggagctatg 1440
ttccctgggt tcttgggagc agcagggaaagc actatggcg cacggtcact gacgctgacg 1500
gtacaggccca gacaattattt gtctggtata gtagcaacagc agaacaattt gctgagagct 1560
attgaggcgc aacagcatct gttcaactc acagtctggg gcatcaagca gctccaggca 1620
agagtccctgg ctgtggaaag atacctaaag gatcaacagc tccttagggat ttggggttgc 1680
tctggaaaac tcatttgcac cactgtgtt cttggaaatg ctagttggag taataatct 1740
ctggatcaga ttttggaaataa catgacactgg atggagtggg agagagaaat tgacattac 1800
acaaaacttaa tatacacctt aattgaagaa tggcagaacc aacaagaaaaa gaatgaacaa 1860
gaatttattag aattttggataa gtggggcaagt ttgtggaaattt ggttggacat atcaaatgg 1920
ctgtggatataaataatgataa gtagggat tagtaggtt aaggatagtt 1980
tttactgtgc ttcttatagt gaatagagttt aggccggat actcaccatt atcatttcag 2040
acccgctcc cagcccccaag gggacccggc agggcccaag gaatcgaaaga agaagggtgga 2100
gagagagaca gagacagatc cagttccattt gtcgtggat tattagcact catctggac 2160
gatctacgga gcctgtgcct cttcagctac caccggctga gagacttaat cttgatttgc 2220
gcgaggattt tggaacttctt gggacgcagg ggggtggaaag ccctcaagta ttggggaaat 2280
ctccctgcagt attggattca ggaactaaag aatagtgtgtt ttagttttttt tgatgcccata 2340
gctatagcag tagctgaggg gacagatagg attatagaag tagcacaaag aattggtaga 2400
gcttttctcc acataacctag aagaataaga caggggctttt aaaggggctttt gctataa 2457

<210> 33
<211> 1453
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp120.modSF162

<400> 33

gaattcgc ccatggatgc aatgaagaga gggctctgt gtgtgtgct gctgtgtgga 60
 gcagtcttcg ttcccccag cgccgtggag aagctgtgg tgaccgtgta ctacggcg 120
 cccgtgtgga aggaggccac caccacccctg ttctgc cca ggcacgcca ggcctacgac 180
 accgagggtgc acaaactgtg gcccacccac gcctgcgtgc ccacccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccccctgt gcgtgaccct gcaactgcacc aacctaaga aacccacca cacaaggagc 420
 agcaactgga aggagatgga cccggcgag atcaagaact gcacgttcaa ggtgaccacc 480
 agcatccgc acaagatgca gaaggagttac gcccgttct acaagctgga cgtgggtgccc 540
 atcgacaacg acaacaccag ctacaagctg atcaactgc acaaccacgat gatcaccac 600
 gcctgccccca aggtgagctt cgagcccatc cccatccact actgcgcccc cgcggcttc 660
 gcccattctg agtgcaacgca caagaagttc aacggcagcg gcccgtgc acacgtgagc 720
 accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 780
 agcctggccg aggaggccgt ggtgatccgc agcgagaact tcacccgacca cgccaaagacc 840
 atcatcgtgc agctgaaaggaa gaggcgtggag atcaactgc accggcccaaa caacaacacc 900
 cgcaagagca tcaccatcg cccggccgc gccttctacg ccacccggcga catcatcg 960
 gacatccggc agggccactg caacatcgc ggcgagaagt ggaacaacac cctgaagcag 1020
 atcgtgacca agctgcaggc ccagtccggc aacaagacca tcgtgttcaa gcagagcagc 1080
 ggcggccgacc cccgagatctg gatgcacagc ttcaactgc gggcggagtt ctctactgc 1140
 aacagcaccc agctgttcaa cagcaccctgg aacaacacca tcggcccaaa caacaccaac 1200
 ggcaccatca ccctgcccctg cccgatcaag cagatcatca accgctggca ggagggtggc 1260
 aaggccatgt acgccccccc catccgccc gatccctgc gcaagcagcaa catcaccggc 1320
 ctgtgtctg acccgacgg cggcaaggag atcagcaaca ccaccggat ctccggcccc 1380
 ggcggccgacc acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtaag 1440
 atcgagccccc tgg 1453

<210> 34

<211> 1387

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp120.modSF162.delV2

<400> 34

gaattcgc ccatggatgc aatgaagaga gggctctgt gtgtgtgct gctgtgtgga 60
 gcagtcttcg ttcccccag cgccgtggag aagctgtgg tgaccgtgta ctacggcg 120
 cccgtgtgga aggaggccac caccacccctg ttctgc cca ggcacgcca ggcctacgac 180
 accgagggtgc acaaactgtg gcccacccac gcctgcgtgc ccacccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccccctgt gcgtgaccct gcaactgcacc aacctaaga aacccacca cacaaggagc 420
 agcaactgga aggagatgga cccggcgag atcaagaact gcacgttcaa ggtgggtgccc 480
 ggcaagctg acaactgc aaccacgtg atcaccacgg ctgcggccca ggtgagcttc 540
 gagcccatcc ccatccacta ctgcggccccc gccggcttc ccacccctgaa gtgcaacgac 600
 aagaagttca acggcagccg cccctgcacc aacgtgagca cccgtgcagtg caccacggc 660
 atccggccccc tggtgagcac ccacgtgctg ctgacccggc gcctggccga ggaggccgtg 720
 gtgatccgcg gcgagaactt caccgacaac gccaagacca tcacgtgc gctgaaggag 780
 agcgtggaga tcaactgcac cccggcccaac aacaacaccc gcaagagcat caccatcg 840
 cccggccgcg ccttctacgc caccggcgac atcatcgccg acatccgcca ggcggccactgc 900
 aacatcagcg gcgagaagtg gaaacaacacc ctgaaaggcga tcgtgacca gctgcaggcc 960
 cagttccggca acaagacccat cgtgttcaag cagagcagcg gccggccaccc cgagatctg 1020
 atgcacagct tcaactgcgg cggcgagttc ttctactgc acaacacca gctgttcaac 1080
 agcaccctgga acaacacccat cggcccaac aacaccaacg gcaccatcac cctggccctgc 1140

cgcatcaagc agatcatcaa ccgctggcag gaggtggca aggccatgta cgcccccccc 1200
 atccgcggcc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgcgcacggc 1260
 ggcaaggaga tcagcaaacac caccgagatc ttccgccccg gcggcggcga catgcgcgac 1320
 aactggcgcga gcgagctgta caagtacaag gtggtaaga tcgagccccct gggcgtggcc 1380
 cccacca 1387

<210> 35
 <211> 1323
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp120.modSF162.delV1V2

<400> 35
 gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcttcg tttcgcccg cggcgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggagggccac caccacccctg ttctgcccga ggcacgcaccc 180
 accggagggtc acaacgtgtg gcccacccac gcctgcgtgc ccaccgcaccc 240
 gagatcggtc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat cagcgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccctgt gcgtggcgcg cggcaactgca cagaccageg tgatcacccca ggcctgcccc 420
 aagggtgagct tggagccat ccccatccac tactgcgcaccc 480
 aagtgcacacg acaagaagt caacggcagc ggccctgtca ccaacgtgag caccgtgcac 540
 tgcacccacg gcatccgccc cgtgggtgac acccagctgc tgctgaacgg caccgtggcc 600
 gaggaggccg tggtgatccg cagcggaaac ttcaaccgaca acggcaacac catcatcg 660
 cagctgaagg agagcgtggatcaactgc accccgcacca acaacaacac cggcaagagc 720
 atcaccatcg gccccccctg cggccgttctac gcccaccccg acatcatcgcc 780
 caggcccact gcaacatcg cggcggaaag tggaaacaacca ccctgaagca gatcgtgacc 840
 aagctgcagg cccaggcccg caacaagacc atcgtttca agcagagcag cggcggcgcac 900
 cccgagatcg tgatgcacag cttcaactgc ggcggcgagt tcttctactg caacagcacc 960
 cagctgttca acagcacctg gaacaacacc atcggccca acaacacca cggcaccatc 1020
 accctgcctt gccgcatcaa gcagatcatc aaccgtggc aggaggtggg caaggccatg 1080
 tacgcggccccc ccattccgccc ccagatccgc tgcaagcaca acatcacccgg cctgtgtctg 1140
 accccgcacg gcccacccatc gatcggcaac accaccgaga tcttccggccc cggcggcggc 1200
 gacatgcgcg acaactggcg cagcggatctg tacaagtaca aggtggtgaa gatcggagccc 1260
 ctggcgtgg ccccccacca ggcacccaccc ggcacccaccc cggcgtggcgc aacgcgagaa ggcgttaactc 1320
 gag 1323

<210> 36
 <211> 2025
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: gp140.modSF162

<400> 36
 gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcttcg tttcgcccg cggcgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggagggccac caccacccctg ttctgcccga ggcacgcaccc 180
 accggagggtc acaacgtgtg gcccacccac gcctgcgtgc ccaccgcaccc 240
 gagatcggtc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat cagcgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccctgt gcgtgaccct gcaactgcacc aacctgaaga acggccacccaa caccaagagc 420
 agcaactgga aggagatggc cccggccgag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgca acaagatgcg aacggatgtac gccccttca acaagctggc cgtgggtgccc 540
 atcgacaacg acaacacccatc atcaactgca acaccagcgt gatcaccatc 600

gcctgccccca aggtgagctt cgagcccatc cccatccact actgcggcccc cgccggcttc 660
 gccatectga agtgcacacga caagaagttc aacggcagcg gcccctgcac caacgtgagc 720
 acggcgact gcaccacacgg catccggcccc gtggtgagca cccagctgct gctgaacggc 780
 agccctggccg aggaggccgt ggtgatccgc agcgagaact tcaccgacaa cgccaaagacc 840
 atcatcgatgc agctgaaggg gaggcggtgg atcaactgc cccggcccaa caacaacacc 900
 cgcaagagca tcaccatcg ccccgccgc geettctacg ccacccggcga catcatcgcc 960
 gacatccggcc aggcccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
 atcgtgacca agtgcaggg ccagttcgcc aacaagacca tcgtgttcaa gcagagcage 1080
 ggccggcacc cccgagatcg tgcacacgc ttcaactgcg gggcgaggt ttctactgc 1140
 aacagcaccc agctgttcaa cagcacctgg aacaacacca tcggcccaa caacaccaac 1200
 ggcaccatca ccctggccctg cccgatcaag cagatcatca accgctggca ggaggtggc 1260
 aaggccatgt acggccccc categcggc cagatccgt gcagcagcaa catcacccggc 1320
 ctgctgctga cccgcgacgg cggcaaggag atcagcaaca ccaccgagat ttccggcccc 1380
 ggcggccggcg acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtaag 1440
 atcgagcccc tggcggtggc ccccaaccaag gccaagcgc gctgggtgca ggcgagaag 1500
 cgcggccgtga ccctggccgc catgttctgg ggcttctgg ggcggccgg cagcaccatgt 1560
 ggcggccgcga gcctgaccct gaccgtgcag gcccggcgc tgctgagcgg catcggtgc 1620
 cagcagaaca acctgctgcg cccatcgag gcccggcgc acctgctgca gctgaccgtg 1680
 tggggcatca agcagctgcg gggccgcgtg ctggccgtgg agegtctacg gaaggaccag 1740
 cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccacccgc cgtggccctgg 1800
 aacgcccaggt ggagcaacaa gagcctggac cagatctgga acaacatgac ctggatggag 1860
 tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
 aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
 aactggttcg acatcagcaa gtggctgtgg tacatctaac tcgag 2025

<210> 37

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.modsF162.delV2

<400> 37

gaatcgccca ccatggatgc aatgaagaga gggctctgtgtgtgtgtgtgtgg 60
 gcaatcgatcg tttcgccccag cgcggctggag aagctgtgg tgaccgtgtta ctacggcgtg 120
 cccgtgtggg aggaggccac caccacccctg ttctgcgcga ggcacccaa ggcctacgac 180
 acccgagggtgc acaacgtgtg ggcacccac gcctgcgtgc ccaccggcccc caaccccccag 240
 gagatcgatgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat caccgtgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 acccccccgtgt gctgtaccct gcaactgcacc aacctgtggaa acgcacccaa caccacggc 420
 agcaactggaa aggagatggaa cccggccgag atcaagaactt gcaatcgatcg ggtggccgc 480
 ggcaagctgtca tcaactgcac caccagcgtg atcaccgggg cctgcggccaa ggtgagctt 540
 gagcccatcc ccatccacta ctgcggcccc gccggcttcg ccattctgaa gtgcacacgc 600
 aagaagttca acggcagcgg cccctgcacc aacgtgagca cctgtcagtg cacccacggc 660
 atccggccccc tggtgagcacc ccaactgtgtg ctgaacggca gcctggccga ggagggcgtg 720
 gtgtatccgca gcgagaactt caccgacaa gccaagacca tcacgtgtca gctgaaggag 780
 agcgtggaga tcaactgcac cccggcccaac aacaacaccc gcaagagcat caccatcgcc 840
 cccggccgcg cttctacgc caccggcgc acatcgccg acatccgcac ggccactgc 900
 aacatcgacgc gcgagaagtg gaacaacacc ctgaaggcaga tcgtgaccaa gctgcaggcc 960
 cagttcgccca acaagacccat cgtgttcaag cagagcggc gcccggccggc ctagatcgatcg 1020
 atgcacacgt tcaactgcac cggcgagttc ttctactgcac acacccacca gctgttcaac 1080
 agcaccatcgga acaacacccat cccggcccaac aacaccaacgc gcaccatcac cctggccctgc 1140
 cgcacatccaa cccgtggcag gaggtggcga aggccatgt ggcggccccc 1200
 atccggccgcg agatccgtcg caccgacac atcaccggcc tgcgtgtgc cccgcacggc 1260
 ggcaaggaga tcaactgcac caccggatc ttccggcccg gcccggccga catgcgcgcac 1320
 aactggcgcac gcgagctgtca caagtacaag gtggtaaga tcgagccctt gggcgtggcc 1380
 cccaccaagg ccaaggcggc cgtggcgtcag cgcgagaagc ggcggcgtgac cctggccgc 1440

atgttcctgg gtttcctggg cgccgcggc agcaccatgg gcgcggcag cctgaccctg 1500
 accgtgcagg cccgcacgt gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
 gccccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcataa gaagctgcag 1620
 gccccgcgtgc tggccgtggg ggcgttacctg aaggaccagc agctgtggg catctggggc 1680
 tgcagcggca agctgtatctg caccacggcc gtgccttggg acgcacgtgt gagcaacaag 1740
 agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gategacaac 1800
 tacacccaacc ttagtacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
 caggagctgc tggagctggg caagtggggc agcctgtggg actgggttgcg catcagcaag 1920
 tggctgtggt acatctaact cgag 1944

<210> 38

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.modSF162.delV1/V2

<400> 38

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gctgtgtgg 60
 gcagtcttcg tttcgcccag cgccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
 ccccggtgtggg aggaggccac caccacccctg ttctgcgcac ggcacgcacaa ggcttacgac 180
 acccgagggtgc acaacgtgtg ggcacccac gcctgcgtgc ccacccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 accccccctgt gcgtgacccct gcaactgcacc aacctgaaga acgcacccaa caccagagc 420
 agcaactggg aggatgtgg cccggggcag atcaagaact gcagttcaa gttggggcgc 480
 ggcaagctgtg tcaactgcaa caccgtgt atcacccagg ctgcacccaa gttgagcttc 540
 gagcccatcc ccatccacta ctgegcccccc gcccggcttcg ccattctgtaa gtgcacgcac 600
 aagaagttca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacacggc 660
 atccggccccc tggtagac ccaactgtgt ctgaacggca gcctggccga ggagggcgtg 720
 gtgatccgca gcgagaactt caccgacaa gccaagacca tcacgtgtca gctgaaggag 780
 agcgtggaga tcaactgtc acggcccaac aacaacaccc gcaagagcat caccatggc 840
 cccggccgcg ccttctacgc caccggcgc acatcgccg acatccgcac ggccactgc 900
 aacatcagcg gcgagaagtgg aacaacaccc ctgaagcaga tcgtgaccaa gctgcaggcc 960
 cagttcgca acaagacccat cgttcaag cagagcagcg gcccggaccc cgagatctg 1020
 atgcacagct tcaactgtgg cggcgagttc ttctactgtca acagcaccacca gctgttcaac 1080
 agcacctggg acaacacccat cggcccaac aacaccaacg gcaccatcac cctgcccctgc 1140
 cgcacatcaacg agatcatcaa cccgtggcag gaggtggggca aggccatgtaa ccccccccc 1200
 atccggccggcc agatccgtc cagcagcaac atcacggcc tgcgtgtgc acggcagcggc 1260
 ggcaaggaga tcagcaacac caccggagatc ttccggcccg gcccggcga catgcgcgc 1320
 aactggcgcga gcgagctgtaa caagtacaag gtggtaaga tcgagccctt gggcggtggcc 1380
 cccaccaagg ccaagcgcgg cgtgggtgcag cgcgagaagc ggcggcgtgac cctggggcgc 1440
 atgttcctgg gtttcctggg cccggccggc agcaccatgg ggcggcgcag cctgaccctg 1500
 accgtgcagg cccggcaggt gtcggcgc atcgtgcage agcagaacaa cctgtgcgc 1560
 gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcataa gaagctgcag 1620
 gccccgcgtgc tggccgtgg ggcgttacctg aaggaccagc agctgtggg catctggggc 1680
 tgcagcggca agctgtatctg caccacggcc gtgccttggg acgcacgtgt gagcaacaag 1740
 agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gategacaac 1800
 tacacccaacc ttagtacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
 caggagctgc tggagctggg caagtggggc agcctgtggg actgggttgcg catcagcaag 1920
 tggctgtggt acatctaact cgag 1944

<210> 39

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut.modSF162

<400> 39

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcttcg tttcgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccaccctg ttctgcgcga ggcacgccaa ggcctacgac 180
 acccgagggtgc acaacgtgtg gcccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaaccaa catggtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 acccccccgt gcgtgaccct gcactgcacc aacctgaaga acgcccaccaa caccaagagc 420
 agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgca acaagatgca gaaggagtagc gcccctgttca acaagctgga cgtggtgccc 540
 atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcacccag 600
 gcctgccccca aggtgagctt cgagccatc cccatccact actgcgcccc cgccggcttc 660
 gccatccctga agtgcacacga caagaagttc aacggcageg gcccctgcac caacgtgagc 720
 accctgtcagt gcacccacgg catccgcccc gttggtgagca cccagctgct gctgaacggc 780
 agcctggccg aggaggccgt ggtgatccgc agcgagaact tcaccgacaa cggcaagacc 840
 atcatcgtgc agetgaaagg gacgcgtggag atcaactgca cccgcaccaa caacaacacc 900
 cgcaagagac tcaccatcgcc ccccgccgc gccttctacg ccacccggcga catcatcgcc 960
 gacatccgccc agggccactg caacatcage ggcgagaagt ggaacaacac cctgaagcag 1020
 atcgtgacca agctgcaggc ccagtctggc aacaagacca tcgttcaa gcagagcagc 1080
 ggccggcgacc ccgagatcgt gatgcacacgc ttcaactgct gccggcaggtt ctctactgc 1140
 aacagcaccc agctgttcaa cagcacctgg aacaacacca tcggccacaa caacaccaac 1200
 ggcaccatca ccctgcccgg cccgcatacg cagatcatca accgcgtggca ggaggtggc 1260
 aaggccatgt acgccccccc catccgcccc cagatccgttgcagcataccggc 1320
 ctgctgtgta cccgcgacgg cggcaaggag atcagcaaca ccaccgat ctccgcccc 1380
 ggcggccggcg acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtaag 1440
 atcgagccccc tgggctgtgc cccaccaag gccaaggccc gcgtggtgc ggcgcagaag 1500
 agcgcctgtg ccctggccgc catgttctg ggcttctgg gcccgcggc cagcaccatg 1560
 ggcggccgcga gcctgaccct gaccgtgcag gcccgcacgc tgcgtgacgg catcgtcag 1620
 cagcagaaca acctgtcgtgc cgcacatcgag gcccagcgc acctgtgcac gtcgaccgtg 1680
 tggggcatca agcagctgca ggcgcgtg ctggccgtgg agcgctacct gaaggaccag 1740
 cagctgtgg gcatctgggg ctgcagcgcc aagctgatct gcaccaccgc cgtggccctgg 1800
 aacgcccagct ggagcaacaa gacgcgtggc agatctgga acaacatgac ctggatggag 1860
 tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
 aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
 aactggttcg acatcagcaa gtggctgtgg tacatctaactcag 2025

<210> 40

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut.modSF162.delV2

<400> 40

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcttcg tttcgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccaccctg ttctgcgcga ggcacgccaa ggcctacgac 180
 acccgagggtgc acaacgtgtg gcccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaaccaa catggtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 acccccccgt gcgtgaccct gcactgcacc aacctgaaga acgcccaccaa caccaagagc 420
 agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtggccgc 480
 ggcaagctga tcaactgca caccacgttgc atcaccacccaa ggtgagcttc 540

gagcccatcc ccatccacta ctgcccggcc gcccgttcg ccatcctgaa gtgcaacgac 600
 aagaagttca acggcagcgcc cccctgcacc aacgtgagca cctgtcagtg caccacggc 660
 atccggccccgg tggtgacac ccagctgctg ctgaacggca gcctggccga ggaggcgtg 720
 gtatccgca gcgagaacctt caccgacaac gccaaggacca tcatacgta gctgaaggag 780
 agcgtggaga tcaactgcac cccggcccaac aacaacaccc gcaagagcat caccatcgac 840
 cccggccggc cttctacgc caccggcgac atcatcgccg acatccgca gggccactgc 900
 aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcaggcc 960
 cagttcggca acaagacccat cgtgttcaag cagagcagcg gggcgaccc cgagatcg 1020
 atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcacccca gctgttcaac 1080
 agcaccttgg acaacacccat cggccccaac aacaccaacg gcaccatcac cctgcccctgc 1140
 cgcacatcaagg agatcatcaa cccgtggcgag gaggtggcga aggcatgta cggccccccc 1200
 atccggggcc agatccgctg caccgacaac atcaccggcc tgctgtcgac cccgcacggc 1260
 ggcaaggaga tcagcaacac ccccgagatc ttcggccccgg gggcgccgac catgcgcgc 1320
 aactggcgca gcgagctgta caatgacaaag gtggtaaga tccggccccctt gggcggtggcc 1380
 cccccaagg ccaaggccgc cgtgggtcgag cgcgagaaga ggcgcgtgac cctggggccgc 1440
 atgttcttgg gtttcttggg cggccggcc gacccatgg gggccggcag cctgaccctg 1500
 accgtgcagg cccgcacgt gctgagcgcc atcgtaacg acgagaacaa cctgctgcgc 1560
 gccatcgagg cccacgacca cctgtcgac ctgaccgtgt ggggcattaa gcaacgtgcag 1620
 gcccgcgtgc tggccctgg agcgttacctg aaggaccacg agtgcgtggg catctggggc 1680
 tgcagcggca agtgcgtcg caccacccgcgttgcga acgcccacgtg gagcaacaag 1740
 agcctggacc agatctggaa caacatgacc ttggatggagt gggagcgcga gatgcacaac 1800
 tacaccaacc tgcgttacac cctgtacgag gagagccaga accagcagga gaagaacgag 1860
 caggagctgc tggagctgga caagtggcc accctgtgga actgggtcgacatcagcaag 1920
 tggctgtggg acatctaact cgag 1944

<210> 41
 <211> 1836
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.mut.modsF162.delV1/V2

<400> 41
 gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgg 60
 gcagtcttcg tttcgccccag cgccgtggag aagctgtggg tgaccgtgta ctacggcg 120
 cccgtgtgg aaggaggccac caccacccctg ttctgcgcaca ggcacccca ggcctacgac 180
 accgagggtgc acaacgtgtg ggcacccac acgtgtgc ccacccgaccc caacccccc 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgggtgg 300
 cagatgcacg aggacatcat caccgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccccgtt gctgtggcgc cggcaactgc cagaccacg tgatcaccac ggcctgcccc 420
 aagggtgacgt tccggccat ccccatccac tactgcgcac ccggccggctt cgcacatcc 480
 aagtgcacg acaagaagtt caacggcagc ggcacccgtca ccaacgtgag caccgtgcag 540
 tgcacccacg gcatccgccc cgtgggtgac acccagctgc tgctgaacgg caccctggcc 600
 gaggaggccg tgggtatcccg caccgtggaaat ttcacccgaca acgccaagac catcatcg 660
 cagctgaagg agacgtggaa gatcaactgc acccggccca acaacacac ccccaagagc 720
 atcaccatcg gccccggccg cgccttctac gccacccggcg acatcatcg cgcacatccgc 780
 caggccccact gcaacatcg cggcgagaag tggaaacaaca ccctgtggca gatcggtacc 840
 aagctgcagg cccagttcg caacaagacc atcggttca agcagacgac cggccggccgc 900
 ccccgagatcg tgcgttacac cttcaactgc ggcggcgatgt tcttctactg caacacgacc 960
 cagctgttca acagcacctg gaacaacacc atcggttccca acaacacca cggcaccatc 1020
 accctggccct gccgcatcaa gcagatcatc aaccgtggcc gggagggtggg caaggccatg 1080
 tacggccccc ccatccggcg ccagatccgc tgcagcagca acatcaccgg cctgctgtgt 1140
 acccgccacg gcccggccaa gatcgacac accaccgaga tcttccggcc cggccggccgc 1200
 gacatcgccg acaactggcg caccgtggcc tacaagtaca aggtggtaa gatcgagccc 1260
 ctggggcggtgg ccccccacca ggcacccggc cgcgtggcgc agcgcggagaa gagecgccgtg 1320
 accctggccg ccatgttccctt ggggttccctg ggcggccggc gcagcaccat gggcgccccgc 1380
 accctgtggcc tgcgttccctg ggcggccggc gatcgatcgac gatcgacac 1440

```

aacctgtgc gcgcctatcgaa gggccaggcagcacctgtgc agctgaccgt gtggggcata 1500
aagcagctgc aggccccgtgtctggccgtg gagcgctacc tgaaggacca gcagctgctg 1560
ggcatctggg gctgcagcgg caagctgatc tgccaccacccg ccgtgcctg gaacgccagc 1620
tggagcaaca agagcctgga ccagatctgg aacaacatga cctggatggaa gtgggagcgc 1680
gagatcgaca actacaccaa cctgatctac accctgtatcg aggagagccaa gaaccagcag 1740
gagaagaacg agcaggagactgtggagactg gacaagtggg ccagcctgtg gaactggttc 1800
gacatcagca atggctgtgtacatctaa ctgcqg 1836

```

<210> 42

<211> 2025

<212> DNA

<213> Artificial Sequence

2203

<223> Description of Artificial Sequence:

98148-MU57-modSE162

<400> 42

<210> 43

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut7.modSF162.delV2

<400> 43

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
gcagtcttcg tttcgccca gcccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcggca ggcacgccaa ggcctacgac 180
accgagggtgc acaacgtgtg gcccacccac gcctgcgtgc ccacccgaccc caaccccccag 240
gagatcggtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccccgt cgtgaagctg 360
accccccctgt gcgtgaccct gcactgcacc aacctgaaga acgcacccaa caccaagagc 420
agcaactgga aggagatgga cccggggcggag atcaagaact gcagcttcaa gttggggcgc 480
ggcaagctga tcaactgcaa caccagegtg atcaccagg cctggcccaa ggtgagcttc 540
gagcccatcc ccatccacta ctgggggggg gccggcttcg ccatctgaa gtgcaacgac 600
aagaagttca acggcagcgg cccctgcacc aacgtgagca cctgtcactgtg caccacccggc 660
atccggcccg tggtagcac ccagctgtg ctgaacgca gcctggccga ggagggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaagacca tcatcgca gctgaaggag 780
agcgtggaga tcaactgcac ccggggccaaac aacaacaccc gcaagagcat caccatcgcc 840
cccggccggc ctttctacgc caccggcggac atcatcgcc acatccgcca gcccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaaggaga tcgtgacccaa gctgcaggcc 960
cagttcgccca acaagaccat cgtgttcaag cagagcagcg gcccggaccc cgagatcggt 1020
atgcacagct tcaactgccc cggcgagttc ttctactgca acagcaccac gctgttcaac 1080
agcacctgga acaacaccat cggggccaaac aacaccaacg gcaccatcac cctggccctgc 1140
cgccatcaagc agatcatcaa ccgctggcag gaggtggca aggccatgtt cggggggggcc 1200
atccggccggc agatcggtg cagcagcaac atcaccggcc tgctgtgtac cccgcacggc 1260
ggcaaggaga tcageaaccac caccggagatc ttccggcccg gcccggccga catgcgcgc 1320
aactggcgca gcgagctgtt caagtacaag tggtagaaga tggggccctt gggcggtggcc 1380
cccaccaagg ccatcagcg cttgggtcag agcgagaaga gcccgtgtac cttggggcgc 1440
atgttcttgg gcttcttggg cggccggccg acgaccatgg gcccggcag cttgaccctg 1500
accgtgcagg cccgcacgt gctgagcggc atcggtcagc agcagaacaa cctgctgcgc 1560
gccatcgagg cccagcagca cttgtgtcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
gcccgcgtgc tggccgtgg ggcgtacccgt aaggaccagc agctgtggg catctggggc 1680
tgcagcggca agctgtatcg caccacccggc gtggcccttggg acgcacgtt gggggcc 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
tacaccaacc tgatctacac cttgtatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgg acaagtggcc agcctgtgg actggttca gatcagcaag 1920
tggctgttgtt acatctaact cgag 1944

```

<210> 44

<211> 1836

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut7.modSF162.delV1/V2

<400> 44

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
gcagtcttcg tttcgccca gcccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcggca ggcacgccaa ggcctacgac 180
accgagggtgc acaacgtgtg gcccacccac gcctgcgtgc ccacccgaccc caaccccccag 240
gagatcggtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccccgt cgtgaagctg 360
accccccctgt gcgtggccgc cggcaactgc cagaccagcg tggatccacca gcccgtggcc 420
aaggtgacgt tggatccat cccatccac tactgcggcc cccggccgtt cccatccgt 480
aagtgcacg acaagaagtt caacggcggc gggccctgtca ccaacgttggg caccgtgtac 540
tgcacccacg gcatccggcc cttgggttggc acccagctgc tgctgttggg cggccgtggcc 600
gaggagggcg tggatccat cggcggccaa ttcaccggaca acgcacccaa catcatcggt 660

```

cagctgaagg agagcgtgga gatcaactgc acceggccca acaacaacac ccgcaagagc 720
 atcaccatcg gccccggccg cgccttctac gccacggcg acatcatcg cgacatccgc 780
 caggcccact gcaacatcg cggcgagaag tggacaaca ccctgaagca gatcgtgacc 840
 aagctgcagg cccagtcgg caacaagacc atcgtttca agcagagcg cgccggcgac 900
 cccgagatcg tgatgcacag cttaactgc ggcggcgagt tcttctactg caacagcacc 960
 cagctgttca acagcactcg gaacaacacc atcggccca acaacacca cggcaccatc 1020
 accctgcacct gccgcataa gcagatcata aaccgtgtgc aggaggtggg caaggccatg 1080
 taacggccccc ccatccggg ccagatcg tgcagcaga acatcacccg cctgctgtcg 1140
 accccgcgacg gccgcagaagg gattcgaac accacccgaga tcttcgcggg cgccggcgac 1200
 gacatgcgcg acaactggcg caggagctg tacaagtata aggtgtgtaa gategagccc 1260
 ctggggcgtgg ccccccaccaa ggcattcagc agcgtgtgtc agagcgagaa gagcgcgtg 1320
 accctggcgcc ccatgttctt gggcttctg ggcgcggccg gcagcaccat gggcggccgc 1380
 agcctgaccc tgaccgtgca ggcccggccag ctgctgagcg gcatcgtgca gcagcagaac 1440
 aacctgtgc ggcgcattcga ggcccagcg cacctgtgtc agctgaccgt gtggggcattc 1500
 aagcagctgc aggccccgt gctggccgtg gagcgttacc tgaaggacca gcagctgtg 1560
 ggcattctggg gctgcagcgg caagctgatc tgcaccaccc cctgcccctg gaacgcgc 1620
 tggagcaaca agagcctgga ccagatctgg aacaacatga cctggatgga gtggggcg 1680
 gagatcgaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccaggcag 1740
 gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 1800
 gacatcagca agtggctgtg gtacatctaa ctcgag 1836

<210> 45

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut8.modSF162

<400> 45

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gctgtgtgga 60
 gcagtcctcg ttgcggccag cgcgtggag aagctgtgg tgaccgtgtta ctacggcggt 120
 cccgtgtgga aggagggccac caccacccctg ttctgcgcga gcgacgcacca ggccgtacgac 180
 accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccacccgaccc caaccccccac 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgggtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 acccccccgt gcgtgaccct gcactgcacc aacctgaaga acgccaccaa caccaagagc 420
 agcaactggg aggagatggg cgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgc acaagatcga gaaggatgc gcccctgttca acaagctggg cgtgggtgccc 540
 atcgcacaacg acaacacccag ctacaagctg atcaactgcac acaccagcgt gatcaccac 600
 gcctggccca aggtgagctt cgagcccate cccatccact actgcgcggg cgcggcgcttc 660
 gccatcctga agtgcaacga caagaagttc aacggcagcg gcccctgcac caacgtgagc 720
 accgtgcagt gcacccacgg catccggcccc gtggtgagca cccagctgtc getgaacggc 780
 agcctggccg aggagggcgt ggtgatccgc agcgagaact tcacccgacaa cgccaaagacc 840
 atcatcgtgc agctgaagga gagcgtggag atcaactgcac cccgcacccaa caacaacacc 900
 cgcaagagca tcaccatcg cccggccgc gccttctacg ccacccggcga catcatcg 960
 gacatccgcg aggcccaactg caacatcagc ggcgagaagt ggaacaacac cctgaagcg 1020
 atcgtgacca agctgcaggc ccagttcgcc aacaagacca tcgtgttcaa gcagagcagc 1080
 ggcggcgacc ccgagatcgt gatgcacacgc ttcaactgcg gccggcgagtt ctcttactgc 1140
 aacagcaccg agctgttcaa cagcacctgg aacaacacca tcggcccaaa caacaccaac 1200
 ggcacccatca ccctgcctcg cgcgcattcaag cagatcatca accgctggca ggaggtgggc 1260
 aaggccatgt acggcccccac catccggcgc cagatccgt gcagcagacaa catcaccggc 1320
 ctgcgtgtca cccgcacggc cggcaaggag atcagcaaca ccacccgagat ctccggcccc 1380
 ggcggcgccg acatgcgcga caactggcgac agcgagctgt acaagtacaa ggtgggtgaag 1440
 atcgagccccc tggggcgtggc ccccaaccatc gccatcagca gcgtgggtca gagcggagaag 1500
 agcgcgtgtca ccctggggcgcc catgttctcg ggtttctgg gccgcggccgg cagcaccatg 1560
 ggcggccgcga gctgaccct gaccgtgcag gcccgccagc tgctgagcgg cattcgtgcag 1620
 cagcagaaca acctgtcgcc cgccatcagc gcccagcagc acctgctgca gctgaccgtg 1680

tggggcatca agcagctgca ggcccgctg ctggccgtgg agcgctacct gaaggaccag 1740
 cagctgtgg gcatctgggg ctgcagcggc aagctgtatc gcaccacccgc cgtgccctgg 1800
 aacgcacatggagcaaca gagectggac cagatctgga acaacatgac ctggatggag 1860
 tgggagcggc agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
 aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
 aactggttcg acatcagcaa gtggctgtgg tacatctaac tcgag 2025

<210> 46
 <211> 1944
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.mut8.modSF162.delV2

<400> 46
 gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gctgtgtgga 60
 gcagtcttcg ttccgcggc aagctgtgg tgaccgtgtta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcga gcgacgccaa ggcctacgac 180
 accgaggtgc acaacgtgtg ggccacccac gctgcgtgc ccaccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat cggcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 accccccgtg gctgtacccct gcaactgcacc aacctgaaga acgccccccaa caccacggc 420
 agcaactggaa aggagatggaa cccggggcgg 540
 ggcaagctgtcaactgcac caccagcgtg atcaccacccgg cctgcggccaa ggtggagcttc 540
 gagcccatcc ccatccacta ctgcggccccc gccggcttcg ccatctgttgc 600
 aagaagttc acggcggcgg cccctgcacc aacgtgagca cctgtcagtg caccacggc 660
 atccggccccc tggtgagcac ccagctgtgtc ctgaacggca gcctggccga ggagggcgtg 720
 gtgatccgca gcgagaactt caccgacaac gccaagacca tcatcgtgca gctgaaggag 780
 agcgtggaga tcaactgcac cccggcccaac aacaacaccc gcaagagcat caccatggc 840
 cccggccggc cttctacgc caccggcggc atcattcggc acatccgcac gggccactgc 900
 aacatcagcg gcgagaagtgg aacaacaccc ctgaaggcaga tcgtgaccaa gctgcaggcc 960
 cagttcgcca acaagaccat cgtgttcaag cagagcggc gggccggcc ctagatcgtg 1020
 atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcaccata getgttcaac 1080
 agcacctggaa acaacacccat cggcccccac aacaccaacg gcaccatcac cctgcggcc 1140
 cgcacatcaagg agatcatcaa cccgtggcggc gaggtgggca aggccatgtt ccccccccc 1200
 atccggggcc agatccgtg cggcggccggc atcaccggcc tgcgtgtgac cccgcacggc 1260
 ggcaaggaga tcagcaacac caccgagatc ttccggccggc gggccggcga catgcggc 1320
 aactggcgcga gcgagctgtt caagtacaag gtggtaaga tcgagccctt gggcgtggcc 1380
 cccaccatcg ccatcagcggc cgtgggtgcggc agcgagaaga gggccgtgc cctggggcc 1440
 atgttctgtt gtttctgggg cccggccggc agcaccatgg gggccggcggc cctgaccctg 1500
 accgtgcagg cccggccaggc gctgagcggc atcgtgcaggc agcagaacaa cctgtgcggc 1560
 gccatcgagg cccaggcggc cctgtgcggc ctgaccgtgt gggccatcaa gcaatgcggc 1620
 gcccggcgtgc tggccgtggaa gcgctacccgt aaggaccaggc agctgtggg catctggggc 1680
 tgcagcggcga agctgtatcg caccacccggc gtggccgtggaa acggccaggctg gagcaacaag 1740
 agectggacc agatctggaa caacatgacc ttggatggagt gggagcgcga gatcgacaac 1800
 tacaccaacc tcatctacac cctgtatcggc gagagccaga accagcaggaa gaagaacggc 1860
 caggagctgc tggagctggaa caagtggggcc agcctgtggaa actgggttgcga catcagcaag 1920
 tggctgtggt acatctaact cgag 1944

<210> 47
 <211> 1836
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.mut8.modSF162.delV1/V2

<400> 47

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
 gcagtcttcg ttcccccag cgccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcga ggcacgccaa ggccctacgac 180
 accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat caccgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 acccccccgt gctgtgggccc cgccaactgc cagaccagcg tgatcaccca ggccctgcccc 420
 aaggtgagct tcgagccat cccatccac tactgcgcgc cgcggccctt cggccatcctg 480
 aagtgcacg acaagaagtt caacggcagc ggccctgtca ccaacgttag caccgtgcag 540
 tgcacccacg gcatccgccc cgtgtgtgac acccagctgc tgctgaacgg cagcctggcc 600
 gaggagggcg tggtgatccg cagcgagaac ttccaccgaca acgccaagac catcatcgtg 660
 cagctgaagg agagcgtggaa gatcaactgc acccgcccca acaacaacac ccccaagagc 720
 atcaccatcg gccccggcccg cgcccttcac gccaccggcg acatcatcg cgcacatccgc 780
 cagggccact gcaacatcag cgccgagaag tggacaaca ccctgaagca gategtgacc 840
 aagctgcagg cccagttcgg caacaagacc atcgtgttca agcagagcag cggggcgac 900
 cccgagatcg tgatgcacag ttcaactgc ggcggcggagt tcttcactg caacagcacc 960
 cagctgttca acagcacctg gaacaacacc atcggcccca acaacaccaa cggcaccatc 1020
 accctgcacct gcccgcataa gcaatcatac aaccgctgtgc aggagggtggg caaggccatg 1080
 tacggggggcc ccateccgccc ccagatccgc tgccagcagca acatcaccgg cctgtgtctg 1140
 acccggcgacg gggcaagga gatcggcggc accaccggaga tcttccgccc egggggggc 1200
 gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtgtgtaa gatcgagccc 1260
 ctggggcgtg cccccccat cgccatcage agcgtgtgtgc agagcggagc gageggcggtg 1320
 accctggggcg ccatgttctt gggcttcctg ggccggcccg gcagcaccat gggegeccgc 1380
 agecctgaccc tgaccgtgca ggcccggccag ctgtgtggc gcatgtgtca gcagcagaac 1440
 aacctgtgc ggcgcatacg ggcggcggc cacctgtgtc agctgaccgt gtggggcattc 1500
 aagcagctgc aggccccgt gctggccgtg gagcgttacc tgaaggacca gcagctgtg 1560
 ggcatactggg gctgcagcgg caagctgatc tgccaccaccc ccgtggccctg gaacgcccac 1620
 tggagcaaca agggcttgg acaacatgtt cctggatggaa gtggggagcgc 1680
 gagatcgaca actacaccaa cctgtatctac accctgtatcg aggagagcca gaaccagcag 1740
 gagaagaacg agcaggagct gctggagctg gacaagtggg ccagccgtgtg gaactgggtc 1800
 gacatcagca agtggctgtg gtacatctaa ctcggag 1836

<210> 48

<211> 2547

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp160.modsF162

<400> 48

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
 gcagtcttcg ttcccccag cgccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcga ggcacgccaa ggccctacgac 180
 accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat caccgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 acccccccgt gctgtggccct gcactgcacc aacctgaaga acgcccaccaa caccacggc 420
 agcaactggaa aggagatggc cccggggcggc atcaagaact gcagcttca ggtgaccacc 480
 agcatccgcg acaagatgca gaaggagttt gcccgttct acaagctggaa cgtgggtggcc 540
 atcgcacaaacg acaacccacg ctacaagctg atcaactgca acaccggcgt gatccacccag 600
 gcctggccca aggtgagctt cgccggccatc cccatccact actgcggcccc cgccggcttc 660
 gccatccgtg agtgcacgca caagaatgtt aacggcggcgg gcccgtgcac caacgtgagc 720
 accgtgcagt gcaacccacgg cattttttttt gtgggtggc cccagctgtgt gctgaacggc 780
 agccctggccg aggagggcgt ggtgtatccgc agcggagact tcaccggacaa cgccaaagacc 840
 atcatactgtgc agctgtggaa gacgtgtggag atcaactgca cccggcccaaa caacaacacc 900
 cgcaagagca tcaccatcgcc cccggccgcgc gccttcacg ccaccggcga catcatcgcc 960
 gacatccgcg aggccacccgt caacatcagc ggcggagaat ggaacaacac cctgtggcag 1020

WO 00/39302

atcggtacca	agctgcaggc	ccagttcgcc	aacaagacca	tctgttcaa	gcagagcagc	1080
ggcgccgacc	ccgagatcgt	gtgcacagc	ttaactcg	gcggcgagtt	cttctactgc	1140
aacagcaccc	agctgttcaa	cagcacctgg	aacaacacca	tccggcccaa	caacaccaac	1200
ggcaccatca	ccctgcacctg	ccgcatcaag	cagatcatca	accgctggca	ggaggtggc	1260
aaggccatgt	acgcgggggg	cattcgccgc	cagatccgt	gcagcagcaa	catcaccggc	1320
ctgtgtctga	cccgcgacgg	ggcaaggag	atcagcaaca	ccacccgagat	cttccgcccc	1380
ggeggcgccg	acatgcgcga	caactggcgc	agegagctgt	acaagtacaa	ggttgtgaag	1440
atcgagcccc	tgggcgtggc	ccccaccaag	gccaaagcgcc	gcgtggtgc	gcgcgagaag	1500
cgcgccgtga	ccctggggcgc	catgttctgg	ggcttcctgg	gcccgcgg	cagcaccatg	1560
ggcgcccgca	gcctgacccct	gaccgtgcag	gcccgcgc	tgctgagcgg	catcggtcag	1620
cagcagaaca	acctgtgcgc	cgccatcgag	gcccagcage	acctgtgc	gtgcaccgtg	1680
tggggcatca	agcagatgc	ggcccgggtg	ctggccgtgg	agegctaccc	gaaggaccag	1740
cagctgtgg	gcatctgggg	ctgcagcgcc	aagctgatct	gcaccacccgc	cgtgccttgg	1800
aacgcgcagct	ggagcaacaa	gacectggac	cagatctgg	acaacatgac	ctggatggag	1860
tgggagcgccg	agatcgacaa	ctacaccaac	ctgatctaca	ccctgtatcg	ggagagccag	1920
aaccaggcagg	agaagaacga	gcaggagctg	ctggagctgg	acaagtgggc	cagctgtgg	1980
aactggtgc	acatcagcaa	gtggctgtg	tacataaga	tcttcatcat	gatcggtggc	2040
ggcctgggtgg	gcctgcgc	cgttttacc	gtgtcgac	tctgtgaaccg	cgtgcgcagg	2100
ggctacagcc	ccctgagctt	ccagacccgc	ttccccccccc	cccgccggccc	cgaccggcccc	2160
gagggcatcg	aggaggaggg	ccgcgcgc	gaccgcgacc	gcagcagccc	cctggtgac	2220
ggcctgtgg	ccctgatctg	ggacgacctg	cgcagctgt	gcctgttca	ctaccacccgc	2280
ctgcgcgacc	tgatcctgat	cgccgcggc	atcgtggagc	tgcgtggccg	ccggggctgg	2340
gaggccctga	agtactgggg	caacctgtcg	cagtactgg	tccaggagct	gaagaacagc	2400
ggcgtagcc	tgttcgacgc	cattgcacatc	gccgtggccg	agggcaccgc	ccgcacatc	2460
gagggtggcc	agcgcacatgg	ccgcgccttc	ctgcacatec	cccgccgcac	ccgcccaggc	2520
ttcgagcgccg	ccctgtgtat	actcgag				2547

<210> 49

<211> 2466

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modsF162.delV2

<400> 49

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtct gctgtgtgga 60
gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcggca gcgcacccaa ggcctacgac 180
acggagggtgc aacaacgtgtg gggcacccac gcctgcgtgc ccaccgaccc caaccccccaag 240
gagatcgtgc tggagaacgt gaccgagaaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagectgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccctgt gcgtgacccct gcaactgcacc aacctgaaga acgcaccaa caccacagac 420
agcaactgga aggagatggg ccggggcgag atcaagaact gcagettcaa gttgggcgc 480
ggcaagctga tcaactgcaa caccagcggt atcacccagg cctggggccaa ggtgagcttc 540
gagcccatcc ccatteccacta ctggeeeeeeee gccggcttcg ccatctgaa gtgcacacgac 600
aagaagtta aecggcagegg cccctgcacc aacgtgagca cctgtcgatg caccacacggc 660
atccggcccg tggtagcac ccagctgtg ctgaacggca gcctggccga ggagggcgtg 720
gtgatccgca gggagaactt caccgacaaac gccaagagcca tcatctgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccggcccaac aacaacaccc gcaagagcat caccatcgcc 840
cccgcccgcg ctttctacgc caccggcgac atcatcgccg acatccgcaca gccccactgc 900
aacatcagcg gcgagaagtg gaacaacaccc ctgaagcaga tcgtgaccaa gctgcaggcc 960
cagttcgca acaagacccat cgtgttcaag cagagcagcg gggcgaccc cgagatcggt 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgtca acagcacccaa gctgttcaac 1080
agcacctgga acaacacccat ccggcccaac aacaccaacg gcaccatcac cctggccctgc 1140
cgcatcaagc agatcatcaa ccgtggcag gaggtggca aggccatgtta cgggggggggg 1200
atcccgccggcc agatccgtg cagcagcaac atcaccggcc tgctgtgac cccggacggc 1260
ggcaaggaga tcagcaacac caccgagatc tteccggcccg gccggcgccga catcgccgac 1320

aactggcgca gcgagctgt acaagtacaag gtgggtgaaga tcgagccccct gggcggtggcc 1380
cccaccaagg ccaaggcgccg cggtggtcag cgccgagaagc gcgcgcgtac cctggggcgcc 1440
atgttctctgg gtttctctgg cgccgcggc agcaccatgg gcgcgcgcag cctgaccctg 1500
accgtgcagg cccgcacgt gtgcgcggc atcgatgcagc agcagaacaa cctgcgcgc 1560
gccatcgagg cccagcagca cctgcgcagc ctgaccgtgt ggggcataa gcagctgcag 1620
gcccgcgtgc tggccgtggc gcgtacatcg aaggaccgc agctgtggg catctggggc 1680
tgcagcggca agctgatctg caccacccgc gtgcgcgtgg acgcacgtg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tgatggat gggagcgcga gategacaaac 1800
tacaccaacc tcatctacac cctgcgcagc gagagccaga accagcggaa gaagaacgcag 1860
caggagctgc tggagctggc caagtggcc acgcctgtgg actggttcga catcagcaag 1920
tggctgtggg acatcaagat cttcatcatg atcgatggcc gcctgtggg cctgcgcac 1980
gtgttcaccg tgctgagcat cgtgaaccgc gtgcgcagg gctacagccc cctgagcttc 2040
cagaccgcgt tccccggccc cggcgccccg gaccgcggcc agggcatcga ggaggaggggc 2100
ggcgagcgcg accgcgcaccc cggcgccccg ctggatgcacg gcctgcgtgc cctgatctgg 2160
gacgacccgtc gcagccgtgc cctgttccgc taccacccgc tgccgcaccc gatcctgtac 2220
gcccggccca tctgtggagct gctggccgc cggggctggg aggcctgaa gtaactggggc 2280
aacctgtgc agtactggat ccaggagctg aagaacagcg cctgagccgt gtcgacgc 2340
atcgccatcg ccgtggccga gggcacccgc cgcacatcg aggtggccca ggcacatcgcc 2400
cggcccttc tgcacatccc cggccgcac cggccaggct tcgagcgcgc cctgatgtaa 2460
ctcgag 2466

<210> 50
<211> 2358

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gpl60.modSF162.delV1/V2

<400> 50

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gctgtgtgg 60
gcagtcttcg ttctcgccag cgccgtggag aagctgtgg tgaccgtgtt ctacggcggt 120
cccgtgtggc aggaggccac caccacccgt ttctcgccca gcgcgcacaa ggccatcgac 180
accgagggtgc acaacgtgtg ggcacccac gcctgcgtgc ccacccgcac caaccccccag 240
gagatcgatgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatcgacg aggacatcat cggctgtgg gaccagagcc tgaacccctg cgtgaagctg 360
acccttcgtg cggcgacatgc cggccacatgc tgatcaccctt ggcctggccc 420
aaggtagct tcgacccat cccatccac tactgcgc cccggccgtt cgcacatccgt 480
aagtgcacac acaagaacgtt caacggcagc ggcctgtgc ccaacgtgag caccgtgcag 540
tgcacccacg gcatccgccc cgtgtgtggc acccagctgc tgctgaaacgg cggctggcc 600
gaggagggcg tggtgatccg cggcgagaac ttcaacggc acgccaacac catcategt 660
cagctgaagg agagctgtgg gatcaactgc acccgccccca acaacaacac cggcaagagc 720
atcaccatcg gccccggccg cgcccttcac gccacccggc acatcatcg cgcacatccgc 780
caggcccact gcaacatcg cggcgagaag tggaaacaaca ccctgaagca gatcggtgacc 840
aagctgcagg cccagttcg cacaagacc atcgatgttca agcagagcag cggccggccgac 900
cccgagatcg tgatgcacag ttcaactgc gggggcgagt tcttctactg caacagcacc 960
cagctgttca acagcacctg gaacaacacc atcgccccca acaacacccaa cggcaccatc 1020
accctgtccct gccgcacatc gcaacatcg aaccgtgtggc aggagggtgg caaggccatg 1080
tacggggggcc ccatccgccc ccacatccgc tgcaacggc acatcaccgg cctgatgtgt 1140
accgcgcacg cggcgaaagg gatcaactgc accaccgaga tttccggccc cggccggccg 1200
gacatcgccg acaactggcg cggcgacatgc tacaagtaca aggtgggtgaa gategacccc 1260
ctgggggtgg cccccccacaa ggcacccgcg cggcgatgtgc agcgcgacaa gcgcgcgtg 1320
accctgtggc ccatgttccct gggcttctgt ggcgcggccg gcacatccat gggccggccg 1380
agctgtggcc tgacatgc gggccggccg ctgtgtggcgc gcacatcg gcaacacac 1440
aacctgtgc ggcacatcg gggccggccg cacatcgatgc agctgtggcc gtcgatgtgt 1500
aagcagctgc agggccggccgt gtcggccgt ggcgcgtacc tgaaggacca gcaatgtgt 1560
ggcatctggg gtcgacccgcg caagatgtc tgcacccaccc cctgcgcgtg gacgcaccc 1620
tggagcaaca agagcttggc ccacatcg aacaacatcg cctggatggc gtcggccgt 1680

gagatcgaca	actacaccaa	cctgatctac	accctgatcg	aggagagcca	gaaccagcag	1740
gagaagaacg	agcaggagct	gctggagctg	gacaagtggg	ccagcctgtg	gaactgggtc	1800
gacatcagca	agtggctgtg	gtacatcaag	atcttcatca	tgatcgtggg	cggcctgtgt	1860
ggcctgcgca	tcgtgttac	cgtgctgagc	atcgtaacc	gcgtgcgcca	gggctacagc	1920
ccccctgagct	tccagaccccg	tttccccggcc	ccccggggcc	ccgaccggccc	cgagggcatac	1980
gaggaggagg	gccccggcg	cgaccggcgac	cgcagcagcc	ccctgggtgca	cggcctgtgt	2040
gccctgatct	gggacgacct	ggcgagcctg	tgcctgttca	gttaccacccg	cctgcgcgac	2100
ctgatcctga	tcgcgcggcg	catcgatggag	ctgatggggcc	ggccgggtgt	ggaggccctg	2160
aagtactggg	gcaacacctgct	gcagtaactgg	atccaggagc	tgaagaacacag	cggcgtgagc	2220
ctgttcgacg	ccatcgccat	cgccgtggcc	gagggcacccg	accggatcat	cgaggtggcc	2280
cagcgcacatcg	gccgcgcctt	cctgcacatc	ccccggccca	tccggccaggg	cttcgagcgc	2340
ccccctgtgt	aactcgag					2358

<210> 51
<211> 1494
<212> DNA
<213> Human immunodeficiency virus

<400> 51
 acaacagtct tgtgggtcac agtctattat ggggtacctg tggaaaaga agcaaccacc 60
 actctgttt gtgcatcaga tgctaaagca tacaaggcag aggacataa cgtctggct 120
 acacatgcgt gtgtacccac agaccccaac ccacaggaag taaatttaac aaatgtgaca 180
 gaaaattta acatgtggaa aaataaacatg gtggAACAGA tgcatgagga tataatcagt 240
 ttatggatc aaaggctaaa gccatgtgtaa attaaaccc cactctgtgt tactttaaat 300
 tgtaactgata agttgacagg tagtactaat ggcacaaataa gtactagtgg cactaatagt 360
 actagtggca ctaatagtac tagtactaat agtactgtata gttggaaaaa gatgccagaa 420
 ggagaaaataa aaaactgctc tttcaatattt accacaagtg taagagataa agtgcagaaa 480
 gaatatttc tctttatataa acttgatgtaa gtaccaatag ataatgataa tgcttagctat 540
 agattgataa attgtatatac ctcagtcattt acacaaggct gtccaaaggt atctttgaa 600
 ccaattccca tacattattt tgccccggctt ggttttgcgat ttctaaagtgt taaagataag 660
 aagtcaatg gaacaggacc atgtaaaaat gtcagcacag taccatgcac acatggaaatt 720
 agaccatgt tatcaactca actgtgttta aatggcagtc tagcagaaga agagatagta 780
 ctttagatctg aaaatttcac agacaatgtc aaaaccataa tagtacagct gaatgaatct 840
 gtagaaattt attgtatataag acccaacaat aatacaagaa aaagtataca tataggacca 900
 gggagagcat ttatgtcaatg aggttatata ataggagaca taagacaagc acattgtaaac 960
 attagtaaag caaactggac taacacttta gaacagatag ttggaaaat aagagaaccaa 1020
 tttgggaata ataaaaacaat aatctttat tcatccctcg gaggggaccc agaaaattgtta 1080
 tttcacagtt ttaattgtgg aggggaattt ttcttattgtat atacatcaca actatttaaat 1140
 agtacctgga atattactga agaggttaaat aagactaaag aaaaatgacac tatcataactc 1200
 ccatgcagaa taagacaaat tataaacatg tggcaagaag tagggaaaagc aatgtatgcc 1260
 cctcccatca gaggacaaat taaatgttca tcaaataatc cagggtctgtt attaactaga 1320
 gatgggtggta ctaacaataa taggacgaac gacacccgaga ctttcagacc tgggggagga 1380
 aacatgaagg acaattggag aagtgaatta tataaatata aagttagtaag aattgaacca 1440
 tttaggatgttag cacccaccca ggcaaaagaga agagtggtgc aaagagagaa aaga 1494

<210> 52
<211> 2007
<212> DNA
<213> Human immunodeficiency virus

<400> 52
acaacagtct tgggttcac agtctattat ggggtacctg tggaaaga agcaaccacc 60
actctgtttt gtgcatacaga tgctaaagca tacaaagcag aggacataa cgtctggct 120
acacatgcct gtgtacccac agaccccaac ccacaggaaag taaaatttaac aaatgtgaca 180
aaaaatttta acatgtggaa aaataaacatg gtggaaacaga tgcatacgat tataatcagt 240
ttatgggatc aaaggctaaa gccatgtgtaa aatattaaaccc cactctgtgt tactttaaat 300
tgtactgata agttgacagg tagtactaat ggcacaaaata gtactgtgg cactaatagt 360
actagtggca ctaatagtac tagtactaat agtactgata gttggaaaaa gatgccagaa 420
ggagaaaataa aaaactgttc ttcaatatac accacaagtg taagagataa agtgcagaaa 480

gaatatttctc	tctttctataa	acttgatgt	gtaccaatag	ataatgataa	tgcttagctat	540
agattgtataa	attgttaatac	ctcagtcat	acacaagect	gtccaaagg	atctttgaa	600
ccaattcccc	tacattattg	tgccccggct	ggtttgcga	ttctaaagt	taaagataag	660
aagttcaatg	gaacaggacc	atgtaaaaat	gtcagcacag	tacaatgcac	acatggatt	720
agaccagtag	tatcaactca	actgctgtt	aatggcagtc	tagcagaaga	agagatagta	780
cttagatctg	aaaatttcac	agacaatgt	aaaaccataa	tagtacagct	gaatgaatct	840
gtagaaaatta	attgtataag	acccaacaat	aatacaagaa	aaagtataca	tataggacca	900
gggagagacat	tttatgcaac	aggtgatata	ataggagaca	taagacaagc	acattgtaac	960
attagtaaag	caaactggac	taacacttta	gAACAGATAG	ttggaaaatt	aagagaacaa	1020
tttgggaata	ataaaaacaat	atctttaat	tcatcctcag	gaggggaccc	agaaaattgt	1080
tttcacagtt	ttaattgtgg	aggggaattt	ttctatgt	atacatcaca	actatTTAAT	1140
agtacctgga	atattactga	agaggtaaat	aagactaaag	aaaatgcac	tatcataactc	1200
ccatgcagaa	taagacaaaat	tataaacat	tggcaagaag	tagaaaagc	aatgttatgcc	1260
cctccatca	gaggacaaaat	taaatgttca	tcaaattatta	cagggtctgt	attaactaga	1320
gatggtggt	ctaacaataa	taggacgaac	gacaccggaga	ccttcagacc	tgggggagga	1380
aacatgaagg	acaattggag	aagtgttata	tataaataata	aagttagtaag	aattgaacca	1440
tttaggatgt	caccaccca	ggcaaagaga	agagtgtgc	aaagagagaa	aagagcagt	1500
ggacttagag	ctttttcat	tggtttctt	gggacgcag	gaagcactat	gggcgcagcg	1560
tcagtgcacgc	tgacgttaca	ggccagacaa	ttattgtctg	gtatagtgc	acagcagaac	1620
aatttgcgt	gagctattga	ggcgcaacag	catctgttgc	aactcacgg	ctggggcattc	1680
aaacagctcc	aggcaagaat	cctggctgt	gaaagatacc	taaaggatca	acagctctt	1740
gggattttgg	gttgcctgg	aaaactcatt	tgcaccaacta	ctgtgcctt	gaactctagt	1800
tggagtaata	aatctctgac	ttagatttg	gataatatga	cctggatgg	gtggggaaaga	1860
gaaattggca	attatacagg	cttaatatac	aatttaattt	aaatagcaca	aaaccagcaa	1920
aaaaagaatg	aacaagaatt	attggattt	gacaagtggg	caagttt	gaattggtt	1980
gatataacaa	actggctgt	gtatata				2007

<210> 53
<211> 2532
<212> DNA
<213> Human immunodeficiency virus

```

<400> 53
acaacagtct tgtgggtcac agtctattat ggggtacctg tggtaaaaga agcaaccacc 60
actctgttt gtgcatcaga tgctaaagca tacaaagcag aggcacataa cgtctggct 120
acacatgect gtgtacccac agaccccaac ccacaggaaag taaaatttaac aaatgtgaca 180
aaaaattttt acatgtggaa aaataaacatg gtggacacaga tgcacgggata tataatcagt 240
ttatggatc aaaggctaaa gccatgtgtaa attaaaccc cactctgtgt tactttaaat 300
tgtactgata agttgacagg tagtactaat ggcacaaata gtactgtgg cactaatagt 360
actagtggca ctaatagtac tagtactaat agtactgtata gttggggaaa gatgccagaa 420
ggagaaataa aaaactgctc tttcaatatac accacaagtg taagagataa agtgcagaaa 480
aatatttctc tctttatataa acttgatgtaa gtaccaatag ataatgataa tgctagctat 540
agattgataa attgtatatac cttagtcatt acacaaggct gtccaaaggt atctttgaa 600
ccaattccca tacattattt tgccccggct ggttttgca ttctaaagtg taaagataag 660
aagtcaatg gaacaggacc atgtaaaaat gtcagcacag tacaatgcac acatggaaat 720
agaccatgt tatcaactca actgctgtta aatggcagtc tagcagaaga agagatagta 780
cttagatctg aaaatttcac agacaatgtc aaaaccataa tagtacagct gaatgaatct 840
gtagaaaaattt attgtatataag acccaacaata aatacaagaa aaagtataca tataggacca 900
gggagagacat tttatgcacaa aggtgatata ataggagaca taagacaagg acattgttaac 960
attagtaaag caaactggac taacacttta gaacagatag ttggaaaaattt aagagaacaa 1020
tttgggataataaaacaat aatctttat tcatcctcag gaggggagcc agaaaattgtt 1080
tttcacagtt ttaattgtgg aggggaattt ttctattgtt atacatcaca actatttaat 1140
agtaccttggatattactga agaggttaaat aagactaaag aaaaatgacac tatcataactc 1200
ccatgcagaa taagacaaat tataaacatg tggcaagaag tagaaaaagc aatgttatgcc 1260
cctccatca gaggacaat taaatgttca tcaaataat cagggctgtt attaactaga 1320
gatgggtgtt ctaacaataa taggacgaaac gacacccgaga ctttcagacc tgggggagga 1380
aacatgaagg acaatttggag aagtgaatta tataaaatata aagttagtaag aattgaacca 1440
ttaggatgtt caccacccca ggcaaaagaga agagtgtgtc aaagagagaa aagagcagt 1500
ggacttaggat ctttgcattt tgggttcttgg gggcggcggcga 1560

```

tcagtgcgc tgacggtaca ggccagacaa ttattgtctg gtatagtgca acagcagaac 1620
 aatttgcgtga gagctattga ggcgcaacag catctgttgc aactcacggc ctggggcattc 1680
 aaacagctcc aggcaagaat cctggctgtg gaaagatacc taaaggatca acagcttcta 1740
 gggattttggg gttgctctgg aaaactcatt tgcaaccacta ctgtgccttg gaactcttagt 1800
 tggagtaata aatctctgac tgagatttg gataatatga cctggatggc gtggggaaaga 1860
 gaaaattggca attatacagg ctaaatatac aatttaatttga aatatacaca aaaccagcaa 1920
 gaaaagaatg aacaagaat attggattt gacaagtggg caagttgtg gaatttggitt 1980
 gatataacaa actggctgtg gtatataaga atattcataa tgatagtagg aggcttgata 2040
 ggtttaagaa tagtttttc tgactttct atagtgaata gagttaggca gggatactca 2100
 ccaatcatcat tgcagacccg cctcccagct cagaggggac ccgcacaggcc cgaaggaaatc 2160
 gaagaagaag gtggagagag agacagagac agatccaatc gattatgtca tggattattt 2220
 gcactcatctt gggacatct gccggagcctg tgcctcttca gctaccaccc cttggagagac 2280
 ttactcttga ttgtagcgag gattgtggaa cttctggac gcaggggggtg ggaagccctc 2340
 aagtatttgtt ggaatctctt gcagtatttg agtcaggagc taaagagtag tgctgttagt 2400
 ttgtttaatg ccacagcaat agcagtagct gaagggacag ataggattat agaaatagta 2460
 caaagaattt ttagagctgt aattcacata cctagaagaa taagacaggg cttggagagg 2520
 gctttactat aa 2532

<210> 54

<211> 1599

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp120.modUS4

<400> 54

gaatttcgcga ccatggatgc aatgaagaga gggctctgtgtgtgtgtgtgg 60
 gcagtcctcg ttctccccag cggcaccacc gtcgtgtgg tgaccgtgtta ctacggcggt 120
 cccgtgttga aggagggccac caccacccctg ttctgcggca ggcacgcacaa ggcttacaag 180
 gcccggggcc acaacgtgtg gcccacccac gcctgcgtgc ccaccgaccc caaccccccac 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catgggtggag 300
 cagatgcatttggaggatcatcat cagccctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 acccccccgtt gctgtaccctt gaaactgcacc gacaagctga cggcagcac caacggcacc 420
 aacagcacca gcccggccaa cagcaccaggc ggcaccaaca gcaccagcac caacagcacc 480
 gacagctggg agaagatgcc cgagggcgag atcaagaact gcagttcaa catcaccacc 540
 aegtgcgcgc acaaggtgca gaaggagttac agcctgttct acaagctgga cgtgggtgccc 600
 atcgacaaacg acaacggccag ctacccgtt atcaactgtca acaccagcgt gatcacccag 660
 gcctggccca aggtgagctt cgagccctt cccatccact actgcggcccc cggccggcttc 720
 gccatcttgcgatgtcaagggca caagaagttt aacggcaccg gcccctgtcaa gaacgtgagc 780
 accgtgcagt gcaacccacgg catccggccccc gtggtgagca cccagctgtct gctgaacggc 840
 aegctggcccg aggaggagat cgtgtgcgc tccggagaact tcacccgacaa cggccaaagac 900
 atcatcggtgc agctgaacgca gtcgggtggag atcaactgtca tccggccccc caacaacacg 960
 cgtaagagca tccacatccgg ccccgccgcgc gccttctac ccacccggca catcatcgcc 1020
 gacatccggcc agggccacttgc caacatccggc aaggccaaact ggaccaacac cctcgagcag 1080
 atcggtggaga agctgcgcgc gcaacttgcgca aacaacaaga ccacatctt caacagcagc 1140
 aegccggccgc accccggat cgtgttccac agcttcaact gcccggccgc gttcttctac 1200
 tgcaacacca gccagctgtt caacacggcc tggaaacatca ccggaggaggt gaacaagacc 1260
 aaggagaacg acaccatcat cctggccctgc cgcatccggcc agatcatcaa catgtggcag 1320
 gaggtggggca aggccatgtt cggccggccccc atccggccgc agatcaagtg cagcagcaat 1380
 attaccggcc tgcgtgtgac cccgcacggc ggcaccaaca acaacccgcac caacgcacacc 1440
 gagaccttcc gccccggccgg cggcaacatgtt aaggacaact ggcgcagcga gctgtacaag 1500
 tacaagggtgg tgcgtgtgac gccccctggc gtggccccc cccaggccaa ggcgcgcgtg 1560
 gtgcagcgcgc agaagcgcta agatatcgga tcctctaga 1599

<210> 55

<211> 1350

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp120.modUS4.del 128-194

<400> 55

```

gaattcggcca ccatggatgc aatgaagaga gggctctgtc gtgtgtgtc gctgtgtgga 60
gcagtcttcg tttcgccca gcccaccacc gtgtgtgtgg tgaccctgtta ctacggcgtg 120
ccccgtgtgga aggaggccac caccaccctg ttctgcgcac ggcacggccaa ggcttacaag 180
ggcgaggccc acaacgtgtc ggcacccac gcctgcgtgc ccaccggccc caaccccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgtag 300
cagatgcatttgggaggacatcat cggctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
accccccctgt gcgtggggc aggaaactgc gagaccageg tgatcaccca ggctgtcccc 420
aaggtagact tcgagccat cccatccac tactgcgcc cgcggcgtt cggccatctg 480
aagtgcagg acaagaagtt caacggcacc ggccctgtca agaacgttag caccgtgcag 540
tgcacccacg gcatccggcc cggtgtgagc acccagctgc tgctgaacgg cagcctggcc 600
gaggaggaga tcgtgtgtcg ctccgagaac ttccacggaca acgcaacac catcatactg 660
cagctgaacg agtccgtgga gatcaactgc atccggccca acaacaacac gcttaagagc 720
atccacatcg gccccggcc cgcccttac gccacccggc acatcatacg cgacatccgc 780
caggcccact gcaacatcag caaggccaa tggaccaaca ccctcgagca gatcgtggag 840
aagctgcccggc agcagttcgca aacaacaag accatcatct tcaacacggc cagcggccgc 900
gaccccgaga tcgtgttcca cagttcaac tgcggccggc agttttcta ctgcaacacc 960
agccagctgt tcaacacggc ctggacatc accggaggagg tgaacaagac caaggagaac 1020
gacaccatca tcctgcctg cccatccgc cagatcatca acatgtggc ggaggtggc 1080
aaggccatgt acggccccc catccgggc cagatcaagt gcagcagcaa tattaccggc 1140
ctgtgtgtca cccggcggc cggcggccaa aacaaccgca ccaacggacac cgagaccc 1200
cgccccggcc gggcaacat gaaggacaac tggcggcggc agctgtacaa gtacaagggt 1260
gtgcgcacatcg agccccctggg cgtggccccc acccaggccca agcggccgcgt ggtgcagcgc 1320
gagaagcgct aagatatcgatcccttaga 1350

```

<210> 56

<211> 2112

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp140.modUS4

<400> 56

```

gaattcggcca ccatggatgc aatgaagaga gggctctgtc gtgtgtgtc gctgtgtgga 60
gcagtcttcg tttcgccca gcccaccacc gtgtgtgtgg tgaccctgtta ctacggcgtg 120
ccccgtgtgga aggaggccac caccaccctg ttctgcgcac ggcacggccaa ggcttacaag 180
ggcgaggccc acaacgtgtc ggcacccac gcctgcgtgc ccaccggccc caaccccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgtag 300
cagatgcatttgggaggacatcat cggctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
accccccctgt gcgtggggc aggaaactgc gagaccageg tgatcaccca ggctgtcccc 420
aacagcaccacccggcaccac cggcaccacc gcaagctgac ccggcggcacc caacaggcacc 480
gacagctggg agaagatgcc cgaggccggatcaacact gcaacgttcaaa catcaccacc 540
agcgtgcgcg acaagggtgca gaaggaggatc agcctgttca acacactgca acaccggcgt gatcaccacc 600
atcgacaacg acaacggccacg ctaccggccatc atcaactgca acaccggcgt gatcaccacc 660
gcctgccttca acgtgagctt cggccatc cccatccact actgcggccccc cgccggcc 720
gccatcctgca agtgcaaggg aacggcaccgg gcccctgcaaa gaacgtgagc 780
accgtgcagt gcaacccacgg catccggccccc gtgtgtgagca cccagctgtc gctgaacggc 840
agccctggccg aggaggagat cgtgtgtgcgc tccgagaact tcaccggacaa cgccaaagacc 900
atcatctgtgc agctgaaacggatc gtccgtggag atcaactgca tccggcccca caacaacacg 960
cgtaagagac tccacatcgcc cccggccgc gccttctacg ccaccggcga catcatcgcc 1020
gacatccggc accggccactc caacatcgac aaggccaaact ggaccaacac cctcgagcag 1080
atcggtggaga agctgcccggc gcaacgttccgc aacaacaaga ccacatctt caacaggcgc 1140
agcggccggcc accggccactc cgtgttccac agcttcaact gccggccgcga gtctttctac 1200
tgcaacaccca gccagctgtt caacaggcacc tggacacatca ccggaggaggt gaacaagacc 1260

```

aaggagaacg	acaccatcat	cctggccctgc	cgcacatccgcc	agatcatcaa	catgtggcag	1320
gagggtggca	aggccatgt	cggccccccc	atcccgcc	agatcaagt	cagcagcaat	1380
attaccggcc	tgtctgtgac	ccgcgacggc	ggcacaaca	acaacccgcac	caacgcacacc	1440
gagaccttcc	gccccggcgg	cgcaacatg	aaggacaact	ggcgcagega	gctgtacaag	1500
tacaagggtgg	tgcgcacatg	gccccctgggc	gtggccccc	cccaaggccaa	gcccggcggt	1560
gtgcagcgcg	agaagcgcgc	cgtggggctg	ggcgcctgt	tcatcggtt	cctggggcgcc	1620
gccgggagca	ccatggggcgc	cgcctccgt	accctgaccg	tgcaggcccg	ccagctgtg	1680
agcggcatcg	tgcagcagca	gaacaacctg	ctgcgcgca	tgcaggccca	gcagcacctg	1740
ctgcagctga	ccgtgtgggg	catcaagcag	ctgcaggccc	gcatectggc	cgtggagcgc	1800
tacctgaagg	accaggcagct	gctgggcattc	tggggctgca	gcccgaagct	gatctgcacc	1860
accaccgtgc	cctggAACAG	cagctggagc	aacaagagcc	tgaccgagat	ctgggacaac	1920
atgaccttga	tggagtggga	gcgcgagatc	ggcaactaca	ccggcctgtat	ctacaacctg	1980
atcgagatcg	cccagaacca	gcaggagaag	aacgagcagg	agctgctgga	gctggacaag	2040
tggggccagcc	tgtggAACTG	gttcgacattc	accaactggc	tgtggtatcat	ctaagatatac	2100
ggatccctta	ga					2112

<210> 57
<211> 2112

<212> DNA

<213> Art:

Journal of International Business

2232

(225) Description of Artificial Sequence:
gp140, mut, modus4

gp140.muc.musos4

<400> 57

gaattcggca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
gcagtcttcg tttcgccccag cgccaccacc cgtctgtgg tgaccctgtta ctacggcggt 120
cccgtgtgga aggaggccac caccacectg ttctcgccca ggcacgccaa ggcttacaag 180
gcccggggcc acaacgtgtc gcccaccacc gctgtgtgc ccacccgacc caaccccccag 240
gagggtgaacc tgaccaacgt gaccggaaac ttcaacatgt ggaagaacaa catggtgag 300
cagatgcatttggagacatc cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccgtg gcgtgaccct gaactgcacc gacaagctga ccggcagcac caacggcc 420
aacagcacca gccggcacca cagcaccaggc ggcaccaaca gcaccagcac acacagcacc 480
gacagctggg agaagatgcc cgagggcgag atcaagaact gcaaccttcaa catcaccacc 540
agcgtgcgcg acaagggtgca gaaggagttac agcctttct acaagctgga cgtgggtgccc 600
atcgacaacg acaacgcacg ctaccgcctg atcaactgca acaccagegt gatcaccagg 660
gcctgccccca aggtgagett cgagcccatc cccatccact actgcgcccccc cgccggcttc 720
gcccatacgtga agtgcacaggaa caagaagtcc aacggcaccc gcccctgcaa gaacgtgagc 780
accgtgcagt gcaccacgg catccggcccc gttgtgagca cccagctgtc gctgaacggc 840
agccctggccg aggaggagat cgtctgtgcgc tccgagaact tcacccgacaa cgccaagagcc 900
atcatcgatc agctgcacgg gttccgtggag atcaactgca tccggcccaa caacaacacg 960
cgtaagagca tccacatcgg ccccgccgcg gccttctacg ccacccggcga catcatcg 1020
gacatccggcc aggcccactg caacatcagc aaggccaaact ggaccaacac cctcgagcag 1080
atcggtggaga agctgcgcga gcagttcgcc aacaacaaga ccatcatttt caacagcagc 1140
agcggcgccg acccccgagat cgtgttccac agttcaact gcggccggcga gttcttctac 1200
tgcaacacca gccagctgtt caacagcacc tggAACATCA ccgaggaggt gaacaagagcc 1260
aaggagaacg acaccatcat cttggccctgc cgcatecgcc agatcatcaa catgtggcag 1320
gaggtgggca aggccatgtt cggggggggcc atccggggcc agatcaagtgc cagcagcaat 1380
attaccggggc tgcgtgtgac cggcgcacggc ggccaccaaca acaacccgcac caacgcacacc 1440
gagaccttcc gccccggccg cggcaacatgg aaggacaact ggccgcacgg gctgtacaaag 1500
tacaagggtgg tgcgcacatcg gccccctgggc gtggccccc cccaggccaa ggcggcgctg 1560
gtgcagcgcc agaagagcgc cgtggggctgt ggcgcctgtt tcattcggtt cttggggcgcc 1620
gcccggggccca ccatgggcgc cggccctccgtt accctgaccg tgcaggccccg ccagctgtg 1680
agcggcatcg tgcgcacatcg gaaacaacctgt ctgcgcgcac tgcaggccccca gcaagcacctg 1740
ctgcagctga cggctgtgggg catcaagcag ctgcaggccccca gcatctgtgc cgtggagcgc 1800
taacctgtggaccaccgcgt gctggggccatc tggggctgca ggggcaagct gatctgcacc 1860
accacccgtgc cctggaaacag cagctggggc aacaagagcc tgaccggatcatc ctggggacaac 1920
atgacccgtggatc tggagttgggaa ggcgcacatcg ggcacactaca ccggccctgtatc ctacaacccgt 1980

atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
 tgggccagcc tgtggaactg gttcgacatc accaactggc tgtggatcat ctaagatatac 2100
 ggatcctcta ga 2112

<210> 58
 <211> 2181
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: gp140TM.modUS4

<400> 58
 gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgct gctgtgtgga 60
 gcagtcttc ttgcggccag cgccaccacc gtgctgtggg tgacctgtta ctacggcgtg 120
 cccgtgtgga aggaggccac caccaccctg ttctgcccga gcgacgcca ggcttacaag 180
 gcccaggccc acaaactgtg gcccaccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatgcattg aggacatcat cagcgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccccgtg gctgtgaccct gaactgcacc gacaagctga cccggcagcac caacggcacc 420
 aacacgacca gccggcaccac cagcaccacgc ggcaccaaca gcaccagcac caacagcacc 480
 gacagctggg agaagatgcc cgaggggcag atcaagaact gcagcttcaa catcaccacc 540
 agcgtgcgcg acaagggtgca gaaggagat agcctgttct acaagctgga cgtggtgccc 600
 atcgacaaccc acaacggccac ctaccgcctg atcaactgca acaccagctg gatcaccacc 660
 gcctgccccca aggtgagctt cgaccctact cccatccact actgcgcccc cccggcgttc 720
 gccatcctgaa gtcgcaagga caagaatgtc aacggcaccg gcccctgcaaa gaacgtgagc 780
 accgtgcagt gcacccacgg catccggcccg gtggtgagac cccagctgt gctgaacggc 840
 agcctggccg aggaggagat cgtgtgcgc tcggagaaact tcaccggacaa cggcaagacc 900
 atcatcgtgc agctgaacga gtccgtggag atcaactgca tcggccccaa caacaacacg 960
 cgtaaagagca tcacatcgg cccggccgc gccttctacg ccacccggcga catcatcgcc 1020
 gacatccgccc aggcccactg caacatcage aaggccaact ggaccaacac cctcgagcag 1080
 atcggtggaga agctgcgcga gcagttcggc aacaacaaga ccatcatctt caacagcagc 1140
 agcggcggcgc accccggagat cgtgttccac agcttcaact gccggccgcga gttttctac 1200
 tgcaacacca gccagctgtt caacagcacc tggAACATCA ccgaggaggt gaacaagacc 1260
 aaggagaacg acaccatcat cctggccctgc cgcacccggc agatcatcaa catgtggcag 1320
 gaggtggggca aggccatgtc cggcccccggc atccgcggcc agatcaagtg cagcggcaat 1380
 attaccggcc tgcgtgetgac cccggcggc ggcaccaaca acaacccgcac caacgacacc 1440
 gagacccccc gccccgggg cggcaacatg aaggacaact ggcgcagcga gctgtacaag 1500
 tacaagggtgg tgcgcacatcga gcccctggc gtggccccca cccaggccaa ggcggccgtg 1560
 gtgcagcgcg agaagcgcgc cgtggccctg ggcgcctgt tcacccgtt cctggggcc 1620
 gcccggggca ccatggccgc cgcctccgtg accctgaccc tgcaggcccc ccagctgtg 1680
 agcggcgcac tgcagcagca gaacaacctg tgcgcgcga tggaggcccc gcaacccctg 1740
 ctgcagctga cccgtgtggg catcaaggcg ctgcaggccc gcacccgtt cgtggagcgc 1800
 tacctgtggagg accagcagct gctggccatc tggggctgca gcccggccatc gatctgcacc 1860
 accaccgtgc cctggaaacag cagctggagc aacaagggcc tgaccggat ctgggacaac 1920
 atgacccgttggaa tggagttgggaa gcccggggatc ggcaactaca cccggccctgat ctacaaccc 1980
 atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
 tgggccagcc tgtggaactg gttcgacatc accaactggc tgtggatcat cccatcttc 2100
 atcatgatcg tggccggccct gatccgttgc cgcacccgtt tggccgtgt gaggatcg 2160
 taagatatcg aatcccttag a 2181

<210> 59
 <211> 1818
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.modUS4.delV1/V2

<400> 59

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
 gcagtcttcg ttteccccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcca ggcacgccaa ggcttacaag 180
 gcccaggccc acaacagtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gaggtgaacc tgaccacacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtggggcc 360
 ggcaggcct gcccccaaggt gagcttgcgag cccatccca tccactactg cggcccccgc 420
 ggcttcgcca teetgaaatgt caaggacaag aagttcaacg gcacggggcc ctgcaagaac 480
 gtgagcaccg tgcagtgcac ccacggcata cgccccgtgg tgacgacccca gctgctgtg 540
 aacggcagcc tggccgagga ggagatgtg ctgcgtcccg agaacttcaac cgacaacgc 600
 aagaccatca tctgtgcgtt gaaaggatca actgatcccg ccccaacaaac 660
 aacacgcgtt agagcatcca catcgcccccc ggccgcgcct tctacgcccac cggcgacatc 720
 atcggcaca tccggccaggc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
 gagcagatcg tggagaagct ggcgcagcag ttccggcaaca acaagaccat catcttcaac 840
 agcagcageg ggggcgcaccc cgagatgtg ttccacagct tcaactgccc cggcgagtcc 900
 ttctactgca acaccagcca gctgttcaac agcacccctgga acatcacccga ggaggtgaac 960
 aagaccaagg agaacgcac catcatactg ccctgcgcga tccggccagat catcaacatg 1020
 tggcaggagg tggcaaggc catgtacgccc cccccccatcc gggccagat caagtgcagc 1080
 agcaatatta cccgcctgtt gctgaccgc gacggcggca ccaacaacaa cccgcaccaac 1140
 gacaccgaga cttccgcgcgc cggcggcggc aacatgaagg acaactggcg cagcgagctg 1200
 tacaagtaca aggtgggtgcg catcgagccct ctggggctgg cccccccacca ggccaagcg 1260
 cgcgttgtgc agcgcgagaa ggcgcgcgtg ggcctggcg ccctgttcat cggcttcttg 1320
 ggcgcgcgcg ggagcaccat gggcgcgcgc tccgtgaccc tgaccgtgca ggccgcgcag 1380
 ctgctgagcg gcatcggtgca gcaagcggaaac aacatgtgc ggcgcacatcgaa ggcccagcag 1440
 cacctgtgc agctgaccgt gtggggcata aacgcacgtgc aggcccgcat cctggccgtg 1500
 gagcgctacc tgaaggacca gcaactgtgc ggcacatctgg gctgcagcggg caagctgatc 1560
 tgcacccacca cccgtgcctg gaacagcgc tggagcaaca agacgtgc acgatctgg 1620
 gacaacatca cctggatgga gtggggagcgc gagatcgca actacaccgg ctgtatctac 1680
 aacctgatcg agatcgccca gaaccagcag gagaagaacg agcaggagct gctggagctg 1740
 gacaagtggg ccagcctgtg gaactgggttc gacatccca actggctgtg gtacatctaa 1800
 gatatcgat cctctaga 1818

<210> 60

<211> 2031

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.modUS4.delV2

<400> 60

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgga 60
 gcagtcttcg ttteccccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcca ggcacgccaa ggcttacaag 180
 gcccaggccc acaacagtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gaggtgaacc tgaccacacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccccctgt gctgtaccct gacatgcacc gacaagctga cccgcagcacc caacggcacc 420
 aacagcacca gggcaccaa cagcaccaggc ggcaccaaca gcaccacgc caacagcacc 480
 gacagctggg agaagatgcg cgaggcggag atcaagaact gcaacttcaa catggcgcc 540
 ggcgcgcgtg tcaactgca caccagcgtg atcaccctagg cctggcccaa ggtgagctc 600
 gagccccatcc ccatccacta ctgcgcgcgc gccggcttgc ccacccctgaa gtgcaaggac 660
 aagaagtca acggcaccgg cccctgcaag aacgtgagca cccgtgcgtg caccacggc 720
 atcggccccc tggtgagcac ccagcgtgtg ctgaacggca gcctggccga ggaggagatc 780
 gtgctgcgtt ccgagaacctt caccgacaac gccaagacca tcaactgtgca gctgaacgc 840
 tccgtggaga tcaactgcac ccggccgcac aacaacacgc gtaagagcat ccacatcgcc 900
 cccggccgcg ctttctacgc caccggcgac atcatcgcc acatccgcgca ggcccactgc 960

aacatcagca aggccaactg gaccaacacc ctcgagcaga tcgtggagaa gctgcgcgag 1020
 cagttcgccca acaacaagac catcatctt aacagcagca gcggcggcga ccccagatc 1080
 gtgttccaca gttcaactg cgccggcga ttcttctact gcaacaccag ccagctgttc 1140
 aacagcacct ggaacatcac cgaggaggtg aacaagacca aggagaacga caccatcatc 1200
 ctgeccctgcc gcatecccca gatcatcaac atgtggcagg aggtggcaa ggcctatgtac 1260
 gccccccca tccgeggcca gatcaagtgc agcagcaata ttacccgcct gctgctgacc 1320
 cgcgacggcg gcaccaacaa caacccgacc aacgacaccc agacccctcg cccggcggc 1380
 ggcaacatga aggacaactg ggcgcagcag ctgtacaagt acaaggttg ggcgcattc 1440
 cccctggcg tgccccccac ccagggcaag cgccgcgtgg tgcagcgcga gaagcgcgc 1500
 gtgggcctgg ggcgcctgtt catcgctt ctggcgccg cccggagcac catggcgcc 1560
 gcctccgtga ccctgaccgt gcaggccccg cagctgtca ggcgcattcgt gcagcgcag 1620
 aacaacactgc tgcgcgcctt cgaggccccg cagcaccgtc tgcagctgac ctgtggggc 1680
 atcaagcagc tgcaggccccg catcctggcc gtggagcgtt acctgaagga ccagcagctg 1740
 ctgggcattt ggggctgcag cggcaagctg atctgcacca ccacccgtcc ctggaaacagc 1800
 agctggagca acaagagcct gaccgagatc tgggacaaca tgacccgtt ggagtgggag 1860
 cgcgagatcg gcaactacac cggcctgatc tacaacctga tgcagatcgc ccagaaccag 1920
 caggagaaga acgagcagga gctgctggag ctggacaagt gggccagct gtggaaactgg 1980
 ttcgacatca ccaactggct gtggtaatcgt taagatatcg gatectcttag a 2031

<210> 61

<211> 1818

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut.modUS4.delV1/V2

<400> 61

gaatttcgcca ccattggatgc aatgaagaga gggctctgtt gtgtgtgtgt gctgtgtgg 60
 gcagtcttcg tttcgccca cgccaccacc gtgtgtggg tgaccgtgtt ctacggcggt 120
 cccgtgtgg, aggaggccac caccacccgt ttctgcgcac ggcacgcacca ggcttacaag 180
 gcccaggccc acaacgtgtg ggccacccac gctgtgtgc ccacccgaccc caacccccc 240
 gaggtgtacc tgaccaacgt gaccgagaac ttcaacatgtt ggaagaacaa catgggtgg 300
 cagatgtcatg aggacatcat cagccctgtgg gaccagagcc tgaagccctgt cgtggcgcc 360
 ggccaggccct gccccaaagggt gagcttcgag cccatccccca tccactactg cgcggcc 420
 ggcttcgcca tcctgaagtg caaggacaag aagttcaacg gcacccggcc ctgcaagaac 480
 gtgagcaccg tgcagtgcac ccacggcattc cggccgggtgg tgagcacccca gctgtgtgt 540
 aacggcagcc tggccgagga ggagatcgatc tgcgcgtcc agaacttcac cgacaacgc 600
 aagaccatca tcgtgcacgt gaaagtcgtt gtggagatca actgcattccg ccccaacaac 660
 aacacgcgttca agacatccatc catggccccg ggcggcgccct tctacgcaccc cggcgacatc 720
 atcggcgaca tccggccaggc ccaactgcac atcagcaagg ccaactggac caacaccctc 780
 gaggcagatcg tggagaagct ggcgcagcgtt tccggcaaca acaagaccat catcttcaac 840
 agcagcagcg gccggcgacc cggatcgatc ttccacacatc tcaactgcgtt cggcgagttc 900
 ttctactgtca acaccagcca gctgtcaac agcaccctggaa acatcaccga ggaggtaac 960
 aagaccaagg agaacgacatc catcatccgtt ccctgcggca tccggccatg catcaacatg 1020
 tggcaggagg tggcaaggc catgtacgccc ccccccattcc gggcccgat caagtgcagc 1080
 agcaatatta cccggctgtt gctgaccccg gacggcgccca ccaacaacaa cccgcaccaac 1140
 gacaccggaga ctttccggccc cggcgccggcc aacatgaagg acaactggcg cagcgagctg 1200
 tacaagtatca aggtgggtcg catcgatccctt ctggcggtgg ccccccacccca ggccaagcgc 1260
 cgcgttgtgc agcgcgagaa gagcgcgtg ggcctggcg ccctgttcat cggcttctg 1320
 ggcggccggcg ggagcaccat gggccggcc tccgtgaccct tgaccgtgca ggcggcccg 1380
 ctgtgtgtgc gcatcgatc gcaagcgttca aacctgttgc ggcgcattcga ggcggcccg 1440
 cacctgttgc agctgaccgt gtggggccatc aagcagctgc aggcggccat cctggccgtg 1500
 gagcgcgtacc tgaaggacca gcaatgtgtt ggcacatctgg gctgcagcgg caagctgatc 1560
 tgcacccatca cccgtgcctgtt gacacgcgtt tggagcaaca agacccgttgc cggatctgg 1620
 gacaacatgtt cctggatggt gtggggccgtt gatgtccgtt actacaccgg cctgtatctac 1680
 aacccgtatcg agatcgccca gacccgcgtt gagaagaacg agcaggagatc gctggagctg 1740
 gacaagtgggg ccaggctgtt gacatggatccatc actggcgatc gatcatctaa 1800

gatatcgat cctctaga

1818

<210> 62
<211> 1818
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
gp140.modUS4.del 128-194

<400> 62
gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtc gctgtgtgaa 60
gcagtcttcg tttcgccca cgccaccacc gtgtgtgtgg tgaccgtgtta ctacggcggt 120
cccgtgtgaa aggaggccac caccaccctg ttctgcggca gcgcacgccaa ggcttacaag 180
gccgaggccc acaacgtgtg gcccaccac gcctgcgtgc ccaccgaccc caaccccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcgtg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtggcgcc 360
ggccaggcct gccccaaagggt gagcttcgag cccatccccca tccactactg cgccccccg 420
ggcttcgcca tcctgaagtg caaggacaag aagttcaacg gcaccggcccc ctgcaagaac 480
gtgagcaccc tgcagtgcac ccacggcata cgccccgtgg tgagcacccca gctgtgtgt 540
aacggcagcc tggccgagga ggagatcgtg ctgcgtctcg agaacttcaac cgacaacgccc 600
aagaccatca tcgtgcagct gaacgagatcc gtggagatca actgcatacg ccccaacaac 660
aacacgcgtt agagcatcca catcgcccccc gcgcgcgtt tctacgccc acggcgacatc 720
atcggcgaca tccgcaggc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
gagcagatcg tggagaagct ggcgcagcag ttcggcaaca acaagaccat catcttcaac 840
agcagcagcc gccgcgaccc cgagatcgtg ttccacagct tcaactgcgg cggcgagtcc 900
ttctactgca acaccagccca gctgttcaac agcacctgttga acatcaccga ggaggtgaac 960
aagaccaagg agaagacac catcatectg ccctgcgcga tccgcgcgtt catcaacatc 1020
tggcaggagg tggcaaggc catgtacggc cccccatcc gggcccgat caagtgcagc 1080
agcaatatta cccgcgtgt gctgaccggc gacggcgccca ccaacaacaa cccgcaccaac 1140
gacaccgaga ccttcgcggcc cgccggcgcc aacatgaagg acaactggcg cagcgagctg 1200
tacaagtaca aggtggtgcg catcgagcccc ctggcggtgg ccccccacccca ggccaagcgc 1260
cgcggttgtgc agcgcgagaa gagcgcgcgtg ggcctggcgcc cccgttcat cggcttcctg 1320
ggcgccgcgg ggagcaccat gggcgccgc tccgtgaccc tgaccgtgca ggccccgcgg 1380
ctgctgagcg gcatcgtgca gcagcagaac aacctgtgc gcgcacatcgaa ggccccagcg 1440
cacctgtgc agctgaccgt gtggggcatc aagcagctgc aggcggcat cctggccgtg 1500
gagcgctacc tgaaggacca gcagctgtg ggcacatctgg gctgcagcgg caagctgatc 1560
tgcaccacca ccgtgcctg gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
gacaacatga cctggatgaa gtggggatggc gagatcgca actacaccgg cctgatctac 1680
aacctgtatcg agatcgccca gaaccagcag gagaagaacg agcaggagct gctggagctg 1740
gacaagtggg ccagcctgtg gaactggttc gacatcacca actggctgtg gtacatctaa 1800
gatatcgat cctctaga. 1818

<210> 63
<211> 1863
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
gp140.mut.modUS4.del 128-194

<400> 63
gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtc gctgtgtgaa 60
gcagtcttcg tttcgccca cgccaccacc gtgtgtgtgg tgaccgtgtta ctacggcggt 120
cccgtgtgaa aggaggccac caccaccctg ttctgcggca gcgcacgccaa ggcttacaag 180
gccgaggccc acaacgtgtg gcccaccac gcctgcgtgc ccaccgaccc caaccccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300

cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 aaaaaaaaaaaaaaaaatgt gcgtgggggc agggactgc gagaccagcg tgatcaccaaa ggccctgcccc 420
 aaggtagatc tcgagcccat cccatccac tactgcgc ccgcggctt cgccatctg 480
 aagtgcagg acaagaagtt caacggcacc ggcctgtggca agaacgtgag caccgtgcag 540
 tgcacccacg gtcacccccc ctgggtgagc acccagctgc tgctgaacgg cagcctggcc 600
 gaggaggaga tcgtgtcgctcccgagaac ttccaccgaca acgccaagac catcatctg 660
 cagctgaacg agtccgtggaa gatcaactgc atccggccca acaacaacac gcttaagagc 720
 atccacatcg gccccggcccg cgccttctac gccacccggcg acatcatcg cgacatccgc 780
 caggcccact gcaacatcg caaggccaaac tggaaccaaca ccctcgagca gatctggag 840
 aagctgcgcg agcagttcgcaacaacaag accatcatcg tcaacagcag cagcggccgc 900
 gaccccgaga tcgtgttcca cagttcaac tgccggccggc agttttcta ctgcaacacc 960
 agccagctgt tcaacagcactgcaacatc accggaggagg tgaacaagac caaggagaac 1020
 gacaccatca tcctgcctg cccatccgc cagatcatca acatgtggca ggaggtggc 1080
 aaggccatgt acggccccccc catccggcgc cagatcaagt gcagcagcaa tattaccggc 1140
 ctgtgtctga cccgcgacgg cggcaccacaa aacaaccgca ccaacgacac cgagaccttc 1200
 cggccggccg gcccggccg cggcaccat gaaggacaac tggcgccggc agctgtacaa gtacaagggt 1260
 gtgcgcatcg agccctgggg cgtggccccc acccaggcca agcgcgcgt ggtgcagcgc 1320
 gagaagagcg cctgtggccctt gggggccctg ttcatcggt tcctggccgc cggccgggagc 1380
 accatggccg cccgcctccgt gacccctgacc tgccggccccc gccagctgt gacggccatc 1440
 gtgcagcgcg agaacaacact gtcgtgcgc accggaggccc agcagcacct gctgcagctg 1500
 accgtgtggg gcatcaagca gtcgtggcc cgcattctgg ccgtggagcg ctacactgaag 1560
 gaccagcgc tgctggcat ctggggctgc agcggcaagc tgatctgcac caccaccgtg 1620
 ccctggaaaca gcaactggag caacaagagc ctgaccgaga tctgggacaa catgaccttg 1680
 atggatgtggg agcgcgagat cggcaactac accggccctga tctacaacct gatcgagatc 1740
 gcccagaacc agcaggagaa gaacgagcag gagctgtgg agctggacaa gtggggccagc 1800
 ctgttggaaact gtttcgacat caccaacttgg ctgtgttaca tctaagatata cgatccctct 1860
 aga 1863

<210> 64

<211> 2634

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp160.modUS4

<400> 64

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgg 60
 gcagtcttcg tttcgcccg cgcaccacc accgtgtgggg tgaccgtgtta ctacggcg 120
 cccgtgtggaa agggggccac caccaccctgt ttctgcgc ggcacgcacaa ggcttacaag 180
 gcccggccca acaacgtgtg gcccaccaccc gcttgcgtgc ccaccggccca aaccccccag 240
 gaggtgaacc tgaccaacgt gaccgagaat ttcaacatgt ggaagaacaa catgggtggag 300
 cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 aaaaaaaaaaaaaaaaatgt gcgtggccctt gaaactgcacc gacaagctgt gccgcaccc 420
 aacagcacca gccgcacccaa cagcaccacgc ggcaccaaca gacccggccac 480
 gacagctggg agaagatgcc cgagggcgag atcaagaact gcaacttcaa catcaccacc 540
 agcgtgcgcg acaaggtgca gaaggagttt accgtgttca acaagctgaa cgtgggtggcc 600
 atcgacaacg acaacggccatc ctaccgcctg atcaactgca acaccacgt gatccacccag 660
 gcctgccccca aggtgagctt cgacccatc cccatccact actgcgcggcc cggccggcttc 720
 gccatcttgcg agtgcaggaa caagaagtcc aacggccaccc gcccctgcaaa gacgtgtggc 780
 accgtgcagt gcaacccacgg catccggcccc gtggtgagca cccagctgt gctgaacggc 840
 agcctggccg aggaggagat cgtgtgcgc tccgagaact tcaccgacaa cggccaaagacc 900
 atcatctgtgc agtgcaggaa gtcgtgtggag atcaactgca tccggcccaaa caacaacacg 960
 cgtaaagagca tccacatcgcc cccggccgc gcttctactg ccaccggcga catcatcgcc 1020
 gacatccggcc agggccactg caacatcgc aaggccaaact gacccaaacac cctcgagcag 1080
 atcggtggaga agtgcgcgaa gcaacttgcgca aacaacaaga ccatcatctt caacagcagc 1140
 agcggccggcc accccggat cgtttccac agtgcactt gggccggcga gtttttctac 1200
 tgcacccacca gccagctgtt caacagcacc tggaaacatca cccggaggat gaccaagacc 1260
 aaggagaaacg acaccatcat cctgcctgc cgcattccgc agatcatcaa catgtggcag 1320

gagggtggcga	aggccatgtta	cgcccccccccc	atccggggcc	agatcaagtgc	cagcagcaat	1380
attaccggcc	tgtctgtgac	ccgcgacggc	ggcaccaaaca	acaacccgcac	caacgcacacc	1440
gagaccttcc	gccccggggg	cgccaacatg	aaggacaact	ggcgagcgaa	gctgtacaag	1500
tacaagggtgg	tgcgcatacg	gccccctgggg	gtggccccc	cccaggccaa	gcgcggcggtg	1560
gtgcagcgcg	agaagegcgc	cgtggggctgt	ggcgcctgt	tcatcggttt	cctgggccc	1620
gccccggagca	ccatggggcgc	cgccctccgtg	accctgaccg	tgcaggcccc	ccagctgtg	1680
agcggcatcg	tgcagcagca	gaacaacctg	ctgcgcgcca	tcgaggcccc	gcagcacctg	1740
ctgcagctga	ccgtgtgggg	catcaagcag	ctgcaggccc	gcatacctggc	cgtggagcgc	1800
tacctgaagg	accagcagct	gttggggcattc	tggggctgca	gcggcaagct	gatctgcacc	1860
accaccgtgc	cctggAACAG	cagctggagc	aacaagagcc	tgacccgagat	ctgggacaac	1920
atgaccttgg	tggagtggga	gcgcgagatc	ggcaactaca	ccggcctgtat	ctacaacctg	1980
atcgagatcg	cccagaacca	gcaggagaag	aacgagcagg	agctgtctgg	gctggacaag	2040
tggggcagec	tgtggaatcg	gttcgcacatc	accaacttggc	tgtgttacat	ccgcatactt	2100
atcatgatcg	tggggcgcct	gatecgccctg	cgcacatcggt	tgcgcgtgt	gagcatcg	2160
aaccgcgtgc	gcggggctaa	cagggccatc	agcctgcaga	ccgcctcgcc	cgcccagcgc	2220
ggccccgacc	gccccgggg	catcgaggag	gaggggcg	agcgcgacccg	cgacggcgc	2280
aaccgcctgg	tgcacggcct	gttggggctgt	atctgggacg	acctgcgcag	cctgtcgctg	2340
ttcagctacc	accgcctgcg	cgacactgtct	ctgatcggtt	cccgcatcg	ggagctgtcg	2400
ggccgcgcgc	gctggggagc	cctgaagtac	tggtggaaacc	tgtgcgtat	ctggagccag	2460
gagctgaaga	gcagcgccgt	gacgcgttcc	aacgcacccg	ccatcgccgt	ggccggaggc	2520
accgaccgc	tcatcgagat	cgtgcagcgc	atctcccg	ccgtgatcca	catccccccgc	2580
cgcatccgc	agggccttgg	gcgcgcctgt	ctgtaaagata	tgcggatcctc	taga	2634

<210> 65
<211> 2538

<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp160.modUS4.delV1

<400> 65

gaattcgcca	ccatggatgc	aatgaagaga	gggctctgtct	gtgtgtgtgt	gtgtgtgtgga	60
gcagtcttcg	tttcgccccag	cgccaccacc	gtgtgtgtgg	tgaccgtgt	ctacggcgtg	120
cccggtgtgg	aggagggccac	caccacccctg	ttctgtgtca	gacgacgcca	ggtttacaag	180
gcccaggggcc	acaacgtgt	ggccacccac	gcctgtgtgc	ccacccgaccc	caaccccccag	240
gagggtgaacc	tgaccaacgt	gaccgagaac	ttcaacatgt	ggaagaacaa	catggtgag	300
cagatgcatg	aggacatcat	cagcctgtgg	gaccagagcc	tgaagccctg	cgtgaagctg	360
accccccctgt	gcgtgaccct	gaactgcacc	gacaagctgg	gcgcggcg	cgagatcaag	420
aactgcagct	tcaacatcac	caccagcgtg	cgcgacaagg	tgcagaagga	gtacagccctg	480
ttctacaaga	tggacgtggt	gccccatcgac	aacgacaacg	ccagctaccg	cctgatcaac	540
tgcaacacca	gcgtgatcac	ccaggcctgc	ccccaaaggta	gcttcgagcc	catccccatc	600
cactactgct	ccccccgg	cttcgcacatc	ctgaagtgc	aggacaagaa	gttcaacggc	660
accggccct	gcaagaacgt	gagcacccgtg	cagtgcaccc	acggcatccg	ccccgtgtgt	720
agcaccacgc	tgcgtgtgaa	ccggcagccctg	gcccggagg	agatgtgtct	gwgctccggag	780
aacttcacccg	acaacgccaa	gaccatcatc	gtgcagctga	acgagtcctgt	ggagatcaac	840
tgcatccggc	ccaaacaacaa	cacgcgttaag	agcatccaca	teggccccgg	ccggcgccttc	900
tacgcccac	gacatcat	ccggcgcatac	ccggcagcc	actgcacat	cagcaaggcc	960
aactggacca	acacccctcg	gcagatgtgt	gagaagtc	gwgagcgtt	ccggcaacaa	1020
aagaccatca	tcttacacag	cgacggggc	ggcgcaccccg	agatgtgtt	ccacagcttc	1080
aactgcggcg	gcgagttctt	ctactgcac	accgcacccg	tgttcaacag	caccttggaa	1140
atcaccgagg	aggtgaacaa	gaccaaggag	aacgcacacca	tcatctgtcc	ctggccgcac	1200
cgccagatca	tcaacatgt	gcaggagggt	ggcaagggca	tgtacggccc	ccccatccgc	1260
ggccagatca	agtgcagcag	caatattacc	ggcctgtgtc	tgacccggcga	ccggcggcacc	1320
aacaacaacc	gcaccaacga	cacccgagacc	ttccgcggcc	gcggcggca	catgaaggac	1380
aactggcgca	gcgagctgt	caagtacaag	gtggtgtgc	tcgagccct	gggcgtggcc	1440
cccacccagg	ccaacgcgg	cgtgtgtgc	cgcgagaagc	gcgcgggtgg	cctggggcgcc	1500
ctgttcatcg	gttttcttggg	cgccgggggg	agcaccatgg	gcgcggcctc	cgtgacccctg	1560

```

accgtgcagg cccgccagct gctgagcggc atcgtgcaga agcagaacaa cctgctgcgc 1620
gccatcgagg cccagcagca cctgctcgag ctgaccgtgt ggggcataa gcagctgcag 1680
ccccgcatcc tggccgtgga ggcgtacctg aaggaccagg agtcgtggg catctgggc 1740
tgcagcgcca agctgatctg caccacccacc gtgccttggaa acagcagctg gagcaacaag 1800
agcctgaccc agatctggga caacatgacc tggatggagt gggagcgcga gatcggcaac 1860
tacaccggcc tgatctacaa cctgatcgag atcggccca gaggcaggaa accagcaggaa gaagaacgag 1920
caggagctgc tggagctgga caagtggccc agcctgtggaa actgtttcgaa catcaccaac 1980
tggctgttgtt acatccgcat cttcatcatg atcgtggcg gcctgatcggtt cctgegcatac 2040
gtgttcggcg tgcgtgagcat cgtgaaccgc gtgcgcagg getacagccc catcagccctg 2100
cagaccgcgc tgccgcggca ggcggggccc gaccgcggcc agggcatcgaa ggaggagggc 2160
ggcgagcgcg accgcgaccgc cagcaaccgc ctggtgacag gcctgtggc cctgtatctgg 2220
gacgacctgc gcagccctgtg cctgttcagc taccacccgc tgccgcacct gctgtgtatc 2280
gtggccgcga tcgtggagct gctggggccgc cgccgtggg aggccctgaa gtaactgggtt 2340
aacctgtgc agtactggag ccaggagctg aagagcagcg ccgtgagcc tttcaacgccc 2400
accgcacatcg ccgtggccga gggcaccgac cgcatcatcg agatcgtgcgacgcgtatcttc 2460
cgccgcgtga tccacatccc cccgcgcattt cgccaggggcc tggagcgcgc cctgtgttaa 2520
gatatcggtt cctctaga 2538

```

<210> 66

<211> 2553

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp160_modUS4_delv3

<400> 66

gaattcgcca	ccatggatgc	aatgaagaga	gggctctgct	gtgtgctgct	gctgtgtgga	60
gcagtcttcg	tttcgcccag	cgccaccacc	gtgtgtggg	tgaccgtgta	ctacggcgtg	120
cccggtgtga	aggaggccac	caccacccctg	ttctgtcgcca	gacgcgcca	ggcttacaag	180
gccgaggccc	acaacgtgt	ggccacccac	gcctggctgc	ccactgaccc	caacccccag	240
gaggtgaacc	tgaccaacgt	gaccgagaac	ttcaacatgt	gaaagaacaa	catggtgagg	300
cagatgcatg	aggacatcat	cagcctgtgg	gaccagagcc	tgaaggccctg	cgtgaagctg	360
accccccgtt	gctgtgaccct	gaactgcacc	gacaagctga	ccggcagcac	caacggcacc	420
aacagcacca	gcccaccaa	cagcaccagc	ggcacaaca	gcacaggcac	caacagcacc	480
gacagctggg	agaagatgcc	cgagggcgag	atcaagaact	gcagcttcaa	catggcgcc	540
ggccgcctga	tcaactgcaa	caccagcggt	atcacccagg	cctgccccaa	ggtgagctt	600
gagccccatcc	ccatccacta	ctgccccccc	gcccgttcc	ccatctgtaa	gtgcaaggac	660
aagaagttca	acggcaccgg	ccccctgcaag	aacgtgagca	ccgtgcagtg	cacccacggc	720
atccggcccg	tggtagacac	ccagctgtg	ctgaacggca	gcctggccga	ggaggagatc	780
gtgctgcgt	ccgagaactt	caccgacaac	gccaagacca	tcatctgtca	gctgaacgag	840
tccgtggaga	tcaactgtcat	ccggccccaa	aacaacacgc	gtaagagcat	ccacatcgcc	900
ccggcccg	ccttctacgc	caccggcgac	atcatcgcg	acatccgcca	ggcccactgc	960
aacatcgca	aggccaaactg	gaccacaccc	ctcgagcaga	tctgtggagaa	gtcgccgag	1020
cagttcgcca	acaacaagac	catcatcttc	aacagcagca	gccccggcgca	ccccgagatc	1080
gtgttccaca	gcttcaactg	cgccggcgag	ttctttctact	gcaacaccag	ccagcttttc	1140
aacagcacct	ggaacatcatc	cgaggagggt	aacaagacca	aggagaacga	caccatcatc	1200
ctggccctggcc	gcatccgcca	gatcatcaac	atgtggcagg	agggtggccaa	ggccatgtac	1260
gcccccccca	tccggggcca	gatcaagtgc	agcagacata	ttacccggct	gtgtgtgacc	1320
cgcgacggcg	gcaccaacaa	caacccgcacc	aacgacaccg	agaccttccg	ccccggcgcc	1380
ggcaacatga	aggacaactg	ggcgagcgag	ctgtacaagt	acaagggtgt	gwgcatcgag	1440
ccccctggcg	tggcccccac	ccaggcacaag	cgccggctgg	tgcagcgccg	gaagcgcc	1500
gtgggcctgg	gcccctgtt	catcggttcc	ctggggccgg	ccggggagcac	catggggcgcc	1560
gcctccgtga	ccctgaccgt	gcaggccccgc	cagctgtga	gcggcatcg	gwgacagcag	1620
aacaacctgc	tgcgcgccccat	cgaggccccag	cagcacccgtc	tgcagctgac	cgtgtggggc	1680
atcaagcagc	tgcaggccccg	catcttggcc	gtggagcgct	acctgaaggaa	ccagcagctg	1740
ctgggcattct	ggggctgcag	gggcaagctg	atctgcacca	ccaccgtgccc	ctggaaacagc	1800
agctggagca	acaagagct	gaccgagatc	tgggacaaca	tgacctggat	ggacttggaa	1860

cgcgagatcg	gcaactacac	cgccctgatc	tacaacctga	tcgagatcg	ccagaaccag	1920
caggagaaga	acgagcagga	gctgtggag	ctggacaagt	gggccagcc	gtggaaactgg	1980
ttcgacatca	ccaactggct	gtggtacatc	cgcacatcc	tcatgatcg	gggcggcc	2040
atcgccctgc	gcatcggtt	ccccgtgctg	agcatcgta	accgcgtgc	ccagggctac	2100
agccccatca	gcctcgagac	ccgcctgecc	gcccagcg	gccccgactg	ccccgaggc	2160
atcgaggagg	aggcgccgca	gcgcgaccgc	gacccgacga	accgcctgg	gcacggcc	2220
ctggccctga	tctgggacga	cctgcgcgc	ctgtgcctgt	tcaactacca	ccgcctgc	2280
gacctgtgc	tgatcggtgc	ccgcategtg	gagctgtgg	gcccggcg	ctgggaggcc	2340
ctgaagtact	ggtggAACCT	gctgcagtac	tggagccagg	agctgaagag	cagcggcc	2400
agcctgttca	acgcacccgc	catcgccgtg	gcccgggca	ccgacccat	catcgagatc	2460
gtgcagcgc	tcttcgcgc	cgtgatccac	atccccggcc	gcatccggca	gggcctggag	2520
cgcgcctgc	tgtaaagat	cgatccctct	aqa			2553

<210> 67

<211> 2340

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modus4.delV1/v2

<400> 67

gaattcgcca	ccatggatgc	aatgaagaga	gggctctgct	gtgtgctgt	gtgtgttggaa	60
gcagtctcg	tttcgcccag	cgccaccacc	gtgtgtggg	tgaccgtgt	ctacggcgtg	120
cccggtgtga	aggagggccac	caccaccctg	ttctgcgcca	gcgacgcca	ggttacaag	180
gcccaggccc	acaacgtgt	ggccacccac	gcctgcgtgc	ccaccgaccc	caacccccag	240
gagggtgaacc	tgaccaaactg	gaccgagaac	ttaacatgt	ggaagaacaa	catggtggag	300
cagatgcatg	aggacatcat	cagcctgtgg	gaccagagcc	tgaagccctg	cgtgggcgcc	360
ggccaggcct	gccccaaaggt	gagcttcgag	cccattccca	tccactactg	cgeccccgccc	420
ggcttcgcca	tectgaagt	caaggacaag	aagttcaacg	gcacccggccc	ctgcaagaac	480
gtgagcacccg	tgcagtgcac	ccacggcata	cgecccgtgg	tgagcaccda	gctgctgtg	540
aacggcagcc	tggccgagga	ggagatcgta	ctgcgcgtccg	agaacttcac	cgacaacgccc	600
aagaccatca	tcgtgcagct	gaacgagtc	gtggagatca	actgcatecg	ccccaaacaac	660
aacacgccta	agagcatcca	catcgcccc	ggeegegcet	tctacgcccac	cggcgacatc	720
atcggcgaca	tccggcaggg	ccactgcac	atcageaagg	ccatcgggac	caacacccctt	780
gagcagatcg	tggagaagct	gcccgcag	ttcggcaaca	acaagaccat	cattttcaac	840
agcagcagcg	gccccgaccc	cgagatcgta	ttccacagct	tcaactcg	cggcgatgtc	900
ttctactgca	acaccagcc	gctgttcaac	agcacctgga	acatcaccga	ggaggtgaac	960
aagaccaagg	agaacgacac	catcatcctg	ccctgcggca	tcggccagat	catcaacatg	1020
tggcaggagg	tgggcaggc	catgtacgc	ccccccatcc	gcccgcagat	caagtgcagc	1080
agcaatatta	ccggcctgct	gctgacccgc	gacgggggca	ccaaacaacaa	ccgcaccaac	1140
gacaccgaga	ccttcggccc	ccggcgccgc	aacatgaagg	acaactcg	cagcgagctg	1200
tacaagtaca	aggtgtgtcg	catcgagccc	ctggggcgtgg	ccccccacca	ggccaagcgc	1260
cgcgtgttgc	agcgcgagaa	gcccgcgtg	ggcctggcg	ccctgttcat	cggcttcctg	1320
ggcgccgccc	ggagcaccat	gggcgcgcgc	tccgtgaccc	tgaccgtgca	ggcccgccag	1380
ctgctgagcg	gcatctgtc	gcagcagaac	aacctgtgc	gcgcacatcg	ggcccgacag	1440
cacctgtc	agctgaccgt	gtggggcata	aagcagctgc	aggccccgat	cctggccgtg	1500
gagcgttacc	tgaaggacca	gcagctgtgc	ggcatctggg	gctgcagcg	caagctgtatc	1560
tgcaccacca	ccgtgcctcg	gaacagcagc	tggagaaca	agagcctgac	cgagatctgg	1620
gacaacatga	cctggatgg	gtggggcgc	gagatcgca	actacaccgg	cctgtatctac	1680
aacctgtatcg	agatcccca	haacccgcag	gagaagaacg	agcaggagct	gtggagctg	1740
gacaagtggg	ccagcctgt	gaaactggtc	gacatcaca	acttgcgttg	gtataccctcg	1800
atcttcatca	tgatctgtgg	ccgcctgatc	ggcctgcgc	tcgtgttgc	cgtgctgagc	1860
atcgtgaacc	gctgtgcgca	gggctacagc	cccatcagcc	tgcaaccccg	cctgccccgccc	1920
cagcgcggcc	ccgaccggccc	cgagggcata	gaggaggagg	gcccgcagcg	cgaccgcgcac	1980
cgcagcaacc	gcctgtgtca	ccgcctgtcg	ggcctgatct	ggacgacact	gcccgcgtcg	2040
tgcctgttca	gctaccaccc	cctgcgcgc	ctgctgtga	tcgtggcccg	catcgtggag	2100
ctgctggccc	gcccggcgt	ggaggccctg	aagtactgg	gaaacctgt	gcagtaactgg	2160

agccaggagc tgaagagcag cgccgtgagc ctgttcaacg ccaccgcat cgccgtggcc 2220
 gagggcaccc accgcatcat cgagatcgta cagcgcacatct tccgcgcgt gatccacate 2280
 ccccgccgca tccggcaggg cctggagcgc gcccgtgt aagatatcgat atccctctaga 2340

<210> 68

<211> 2385

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modUS4del 128-194

<400> 68

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gctgtgtgg 60
 gcagtcttcg ttccgcggcc cggcaccacc gtgtgtgtgg tgacctgtta ctacggcg 120
 cccgtgtggaa aggaggccac caccacccctg ttctgcgcga ggcacccaa ggcttacaag 180
 gcccaggccc acaacgtgtg ggcacccac gcctgcgtgc ccacccgaccc caaccccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatgcatg aggacatcat cagegtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccccctgt gcgtgggggc agggaaactgc gagaccagcg tgatcaccca ggctgtcccc 420
 aagggtgagct tcgagcccat ccccatccac tactgcgcgc cgcggcgtt cggccatcc 480
 aagtgcagg acaagaaggta caacggcacc ggcacccctgca agaacgttag caccgtgc 540
 tgacccaccc gcatccggcc cgtgtgtggc acccagctgc tgctgaacgg cagcctggcc 600
 gaggaggaga tcgtgtgtggc ctccgagaac ttacccgaca acgccaagac catcatcg 660
 cagctgaacg agtccgtggaa gataactgc atccggccca acaacaacac gctgtttttt 720
 atccacatcg gcccggcccg cgccttcac gccacccggcg acatcatcg cgacatccgc 780
 caggcccaact gcaacatcg caaggccaaac tggaccaaca ccctcgagca gatcggtgg 840
 aagctgcgcg agcagttcg caacaacaag accatcatct tcaacacgac cagggggc 900
 gaccccgaga tcgtgttcca cagcttcaac tgcggccggcg agtttttcta ctgcaacacc 960
 agccagctgt tcaacacgac ctggAACATC acccgaggagg tgaacaagac caaggagaac 1020
 gacaccatca tcctgcctcg cccatccgc cagatcatca acatgtggca ggaggtggcc 1080
 aaggccatgt acggccccc catccggc cagatcaatg gcagcagcaa tattaccggc 1140
 ctgctgtgtgc cccgcgacgg cggcaccaac aacaaccgca ccaacgcac acgaccccttc 1200
 cggccggccg gccggcaacat gaaggacaac tggcgccagcg agctgtacaa gtacaagggt 1260
 gtgcgcatcg agcccccctgg cgtggccccc acccaggcca agcggccgtt ggtgcagcgc 1320
 gagaagcgccg ccgtggccct gggcgccctg ttcatcggt tccctggccgc cgcggggagc 1380
 accatggcgccg cccctccgt gaccctgacc gtgcaggccc gccagctgt gaggccgatc 1440
 gtgcagcagc agaacaacat gctgcgcgc acccgaggccc agcagcacct gctgcagctg 1500
 accgtgtggg gcatcaagca gctgcaggcc cgcacccctgg ccgtggagcg ctacctgaag 1560
 gaccagcagc tgctggccat ctggggctgc agcggcaagc tgatctgcac caccacccgt 1620
 ccctggaaaca gcaagctggag caacaagagc ctgaccggaga tctgggacaa catgacccctgg 1680
 atggagttggg agcgcgagat cggcaactac accggccctgt tctacaacat gatcgagatc 1740
 gcccagaacc agcaggagaa gaacgagcagc gagctgtgg agctggacaa gtggggccage 1800
 ctgtggaaact gtttcgacat caccaactgg ctgtggatca tccgcacatccatcatgtac 1860
 gtggggccggcc tgatcggtt ggcacatcg ttcggccgtg tgagcatgt gaaacggcg 1920
 cggccaggccat acggcccccatt cggccctgc accggccctgc cggcccgccg cggcccccgc 1980
 cggcccccggagg gcatcgagga ggaggccggc gagcgccgacc ggcacccgcg caaccgcctg 2040
 gtgcacggcc tcgtggccct gatctggac gacctgcgcgac gctgtgtgtt gttcagatc 2100
 cacccgcctgc ggcacccctgt gctgtatcg tcccgccatcg tggagctgtt gggccggcc 2160
 ggctggggagg ccctgaagta ctggggaaac ctgtgtgtt gactggagccca ggagctgaag 2220
 agcagcgcgcg tgacccctgtt caacggccacc gccatcgccg tggccggaggg caccgcaccgc 2280
 atccatcgaga tcgtgcagcgc catccctccgc gcccgtatcc acatccccccg cccatccgc 2340
 caggggccctgg agcgcgcctt getgtaaatgat atccggatcc ctaga 2385

<210> 69

<211> 144

<212> DNA

<213> Human immunodeficiency virus

<400> 69
gacaccatca tcctgccctg cgcgcattccgc cagatcatca acatgtggca ggaggtggc 60
aaggccatgt acgccccccc catcccgccg cagatcaagt gcagcagcaa catcacccggc 120
ctgctgctga cccgcgacgg cgcc 144

<210> 70
<211> 144
<212> DNA
<213> Human immunodeficiency virus

<400> 70
gaaactatca cactccatg cagaataaaa caaattataa acaggtggca ggaagtagga 60
aaagcaatgt atgcccccc catcagagga caaattagat gtcatcaaa tattacagga 120
ctgcttattaa caagagatgg tggt 144

<210> 71
<211> 144
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic Env
US4 common region

<400> 71
gacaccatca tcctgccctg cgcgcattccgc cagatcatca acatgtggca ggaggtggc 60
aaggccatgt acgccccccc catcccgccg cagatcaagt gcagcagcaa catcacccggc 120
ctgctgctga cccgcgacgg cgcc 144

<210> 72
<211> 144
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic Env
SF162 common region

<400> 72
ggcaccatca ccctgccctg cgcgcattcaag cagatcatca accgctggca ggaggtggc 60
aaggccatgt acgccccccc catcccgccg cagatccgt gcagcagcaa catcacccggc 120
ctgctgctga cccgcgacgg cgcc 144

<210> 73
<211> 4766
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
gp160.modUS4.gag.modSF2

<400> 73
gaattcgcgc ccatggatgc aatgaagaga gggctctgtgt gtgtgtgtgt gctgtgtgg 60
gcaggctttcg ttccgcggcag cgccaccacc gtgtgtgtgg tgaccgtgtta ctacggcggt 120
cccggtgtgg agggggccac caccaccctg ttctgcgcgc ggcgcgcggaa ggcttacaag 180
ggcgaggccc acaacgtgtg ggccacccac gctgtgtgc ccaccgcacc caaccccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
cagatgcattt aggacatcat cagccgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360

accccccctgt gcgtgaccct gaactgcacc gacaagctga ccggcagcac caacggcacc 420
 aacagcacca gggcaccaa cagcacccagc ggcaccaaca gcaccagcac caacagcac 480
 gacagctgg agaagatgcc cgaggggcgag atcaagaact gcagcttcaa catcaccacc 540
 agegtgcgc acaagggtca gaaggagttac agcctgttct acaagctgga cgtgggtccc 600
 atcgacaacg acaacccag ctaccgcctg atcaactgca acaccagegt gatcaccacc 660
 gctgtccccca aggtgagctt cgagccatc cccatccact actgccc cgcgggttc 720
 gccatctgtga agtgcaagga caagaagttc aacggcaccg gcccctgcaaa gaacgtgagc 780
 accgtgcagt gcacccacgg catccggccc gtggtgagca cccagctgt gctgaacggc 840
 agcctggccg aggaggagat cgtgtgcgc tccgagaact tcaccgacaa cgccaagacc 900
 atcatcggtc agtgcacca gtcgtggag atcaactgca tccggcccaa caacaacacg 960
 cgtaagagca tccacatcg cccggccgc gccttctacg ccaccggca catcatcgcc 1020
 gacatccccc aggccactt caacatcagc aaggccaaact ggaccaacac cctcgagcag 1080
 atcggtggaga agtgcgcga gcaattccgc aacaacaaga ccatcatctt caacagcagc 1140
 agcggccggcg accccggagat cgtgttccac agcttcaact gcccggcgca gttttctac 1200
 tgcaacaccca gccagctgtt caacagcacc tggAACATCA cccaggagggt gaaaaagacc 1260
 aaggagaacg acaccatcat cctggccctgc cgatccggc agatcatcaa catgtggcag 1320
 gaggtgggca aggccatgtt cgcggggggc atccggggcc agatcaagt gaggcagcaat 1380
 attaccggcc tgcgtgttgc cccggacggc ggcaccaaca acaaccgcac caacgacacc 1440
 gagaccttcc gccccggggc cggcaacatg aaggacaact ggcgcagcga gctgtacaag 1500
 tacaagggtgg tgcgtatcg gccccctggc gtggggggca cccaggccaa gcccggcgtg 1560
 gtgcagcgcg agaaggcgcg cgtggggctg ggcccccgtt tcatcggtt cttggggcgcc 1620
 gccccggagca ccatggggcgc cgcctccgtg accctgaccc tgcaagggccg ccagctgtg 1680
 agcggcatacg tgcagcagca gaaaacactg ctgcgcgc a ctcggggccca gcaagcacctg 1740
 ctgcagctgtt cccgtgtgggg catcaaggag ctgcaggccc gcatcttgc cgtggagcgc 1800
 tacctgttgg accagcagct gctggggatc tggggctgtca gggcaagct gatctgcacc 1860
 accaccgtgc cctggaaacag cagctggagc aacaagagcc tgaccggat ctgggacaac 1920
 atgcacccgttgg tgggtgggg ggcgcagatc ggcacactaca cccgctgt tataaacctg 1980
 atcgagatcg cccagaacca gcaaggagaag aacggcaggc agtgcgttgc gctggacaag 2040
 tggggccagcc tggggactt gttggacatc accaaacttgc tgggttacat cccatctt 2100
 atcatgtatcg tggggccctt gatcgccctg cgcatacggtt tgcgtgtt gacatcggt 2160
 aaccggcgtgc gccagggtca cagccccatc accctgttgc a cccgtgttgc cggccagcgc 2220
 gccccccgacc gccccggggg catcgaggag gaggggggcc aegcgcaccg cgaccgcagc 2280
 aaccggctgg tgcacggccctt gtcgtggccctg atctgggacg acctgcgcag cctgtgttgc 2340
 ttcaagctacc accggctgcg cgcctgttgc ctgcgttgc ggtggggccctt ggcgcgtt 2400
 gggccggccg gctggggaggt cctgttggatc tgggtggacc tgcgtgttgc gtcggagcc 2460
 gagctgttgg gcaaggccgtt ggcctgttgc aacggcaccg ccatcgccgt gggcgaggcc 2520
 accggaccgc tcatcgagat cgtgcagcgc atctccgcg cctgtatcca catccccccgc 2580
 cgcataccgcg aggccctggc ggcgccttgc ctgttggatc tgcgtgttgc tagagaattc 2640
 cggggggggggg cccccccccc ctctccctcc ccccccccttta acgttactgg ccgaagccgc 2700
 ttggataaaag gcccgggtgc gtttgttat atgttatttt ccaccatatt gccgtttttt 2760
 ggcaatgttgc gggccggggaa acctggccctt gtcgtgttgc cgcgttgc gtcgttgc 2820
 tcccccttcg ccaaaaggat gcaaggctgtt gtcgtgttgc ttcgtgttgc tgggttgc 2880
 gaagcttctt gaaagacaaac aacgttgc ttcgtgttgc gtcgtgttgc gtcgtgttgc 2940
 ctcggcgaca ggtgccttgc cggccaaagg ccacgtgttgc aagataacc tgcgtgttgc 3000
 gcacaacccc agtgcacgt tgcgtgttgc atgttgc ttcgtgttgc aaagatcaa atggcttcc 3060
 tcaagctgtat tcaacaaggat gtcgtgttgc gtcgtgttgc gtcgtgttgc ttcgtgttgc 3120
 gatctggggc ctcgggtgcac atgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3180
 gggggggggg accacggggc ctcgggttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3240
 cccggccagc gtcgtgttgc gtcgtgttgc gtcgtgttgc ttcgtgttgc ttcgtgttgc 3300
 cggcgccaaag aagaagtaca agtgcacgt ttcgtgttgc ttcgtgttgc ttcgtgttgc 3360
 cttccgggttgc aaccccccggcc ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3420
 gtcgtgttgc accctgttgc ctcgggttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3480
 caccctgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3540
 gatcgaggag gacgacaaac aacgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3600
 cacccggccaaac agcagccagg ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3660
 gatgggtgcac ctcgggttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3720
 gaaaggccccc ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3780
 ccaggacccgtt aacacgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3840
 gaaaggagacc atcaacggggc ttcgtgttgc ttcgtgttgc ttcgtgttgc ttcgtgttgc 3900

ccccatcgcc	cccggccaga	tgcgcgagcc	ccgcggcagc	gacatcgccg	gcaccaccag	3960
cacctgtcag	gagcagatcg	gctggatgac	caacaacccc	cccatccccg	tggcgagat	4020
ctacaagcgg	tggatcatcc	ttggcctgaa	caagatcgta	cgatgtaca	gccccaccag	4080
catcctggac	atccgccagg	gccccaaagg	gccccttcgc	gactacgtgg	accgcttcta	4140
caagacccctg	cgcgctgagc	aggccagcca	ggacgtgaag	aactggatga	ccgagacccct	4200
gctggtgccag	aaccccaacc	ccgactgcaa	gaccatctcg	aaggctctcg	gccccggcgc	4260
caccctggag	gagatgatga	ccgcctgcca	ggcgctggc	ggccccggcc	acaaggcccc	4320
cgtgctggcc	gagggcatga	gccaggtgac	gaacccggcg	accatcatga	tgcagcgcgg	4380
caacttcgc	aaccagcgg	agaccgtcaa	gtgcttcAAC	tgccgcagg	agggccacac	4440
cgccaggaac	tgccgcgccc	cccgcaagaa	gggctgtgg	cgctgccc	gcgagggcga	4500
ccagatgaag	gactgcaccg	agegcgcaggc	caacttcctg	ggaagatct	gccccagcta	4560
caagggccgc	cccgcaact	tcctgcagag	ccggcccccag	cccacccccc	ccccccgagga	4620
gagcttcgc	ttcggcgagg	agaagaccac	ccccagccag	aaggcaggagc	ccatcgacaa	4680
ggagctgtac	ccccctgacca	gcctgcgcag	cctgttcggc	aacgacccca	gcagccagta	4740
agaattcaga	ctcgagcaag	tctaga				4766

<210> 74

<211> 4689

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modSF162.gag.modSF2

<400> 74

gaattcgcca	ccatggatgc	aatgaagaga	gggctctgtct	gtgtgtgtct	gtctgtgttga	60
gcagtcttcg	tttcggccag	cgcgggtggag	aagctgtggg	tgaccgtgtt	ctacggcggt	120
cccggtgtga	aggaggccac	caccacccctg	ttctgcgcca	gcgacgccaa	ggcctacgac	180
accgagggtgc	acaacgtgtg	ggccacccac	gcctgcgtgc	ccacccgaccc	caacccccag	240
gagatcggtgc	tggagaacgt	gaccgagaac	ttcaacatgt	ggaagaacaa	catggtggag	300
cagatgcacg	aggacatcat	cagecctgtgg	gaccagagcc	tgaagccctg	cgtgaagctg	360
accccccgt	gcgtgaccct	gcactgcacc	aacctgaaga	acgccaccaa	caccaagagc	420
agcaacttgg	aggagatgg	ccggggcgag	atcaagaact	gcagettcaa	ggtgaccacc	480
agcatccgca	acaagatgc	gaaggagtagtac	gcccgttct	acaagtcgg	cgtgggtggcc	540
atcgacaacg	acaacaccag	ctacaagctg	atcaactgtca	acaccagcg	gatcacccag	600
gcctgccc	agggtgagct	cggccccatc	cccattccact	actgcgecccc	cgccggcttc	660
gcacatctga	agtgcAACGA	caagaagttc	aacggcagcg	gcccctgcac	caacgtgagc	720
accgtgcagt	gcacccacgg	catccggcccc	gtggtgagca	cccagctgtct	gctgaacgjc	780
agcctggccg	aggagggcgt	ggtgatccgc	agcgagaact	tcacccgacaa	cgccaagacc	840
atcatcggtc	agctgaagga	gagcgtggag	atcaactgca	cccgccccaa	caacaacacc	900
cgcaagagca	tcaccatcg	ccccggccgc	gccttctacg	ccacccggcg	catcateggc	960
gacatccgc	aggcccactg	caacatcagc	ggcgagaagt	ggaacaacac	cctgaagcag	1020
atcggtgacca	agctgcaggg	ccagttcgcc	aacaagacca	tcgtgttcaa	gcagagcagc	1080
ggcggcgacc	ccgagatcgt	gatgcacagg	ttcaactgcg	gccccggagtt	tttctactgc	1140
aacagcaccc	agctgttcaa	cagcacctgg	aacaacacca	tcggccccaa	caacaccaac	1200
ggcaccatca	ccctggccctg	ccgcatcaag	cagatoatca	accgctggca	ggaggtgggc	1260
aaggccatgt	acgcgggggg	catccggggc	cagatccgtct	gcagcagcaa	catcaccggc	1320
ctgctgtga	cccgcgacgg	cggcaaggag	atcagcaaca	ccacccgagat	cttccggccc	1380
ggcggcgccg	acatgcgcga	caactggcg	agcgagctgt	acaagtacaa	ggtgtgttgaag	1440
atcgagcccc	tggggctggc	cccccaacca	gccaagcg	gctgtgtca	gcgcgagaag	1500
cgccggcgta	ccctggggc	catgttctgg	ggttcttctgg	gccccgggg	cagcaccatcg	1560
ggcggcccgca	gcctgaccct	gaccgtgcag	gccccggcagc	tgctgagcg	catcgtgcag	1620
cagcagaaca	acctgtcg	cgccatcgag	gccccagcgc	acctgtgtca	gctgaccgtg	1680
tggggcatca	agcagctgca	ggccccgcgt	ctggccgtgg	agcgcttact	gaaggaccag	1740
cagctgtgg	gcatctgggg	ctgcagcg	aagctgtatct	gcaccacccgc	cgtggccctgg	1800
aacgcacgt	ggagcaacaa	gaggctggac	cagatctgg	acaacatgac	ctggatggag	1860
tgggagcgcg	agatcgacaa	ctacaccaac	ctgatctaca	ccctgatcg	ggagagccag	1920
aaccagcagg	agaagaacga	gcaggagctg	ctggagctgg	acaagtgggg	cagecctgtgg	1980

aactggttcg acatcagcaa gtggctgtgg tacatcaaga tcttcatcat gatcgtggc 2040
 ggctctggtgg gcctgcgcac cgtttcacc gtgtgagca tcgtgaaccg cgtgcgcac 2100
 ggctacagcc ccctgagctt ccagaccgc ttcccccccc cccgcggccc cgacccccc 2160
 gagggcatcg aggaggaggg cggcgcgc gaccgcgacc gcagcagccc cctggtgac 2220
 ggctctgtgg ccctgatctg ggacgacctg cgcagctgt gcctgtttag ctaccaccgc 2280
 ctgcgcgacc tgatctgtat cggcggccc atcgtggagc tgctggggccg cccgggctgg 2340
 gagggccctga agtactgggg caacctgtc cagtaactgga tccaggagct gaagaacagc 2400
 gcccgtgagcc ttttcacgc categccatc gccgtggccg agggcaccga cccgcacatc 2460
 gaggtggccc agcgcacatcg cccgcgcctc ctgcacatcc cccggcgc cccggcaggc 2520
 ttcgagcgcg ccctgtgtat actcgagcaa gtcttagagaa ttccggccccc cccccccccc 2580
 cccctctccc tccccccccc tccacgttatc tggcgaagc cgcttggaaat aaggccgggt 2640
 tgcgtttgtc tatatgttat ttccacatc attggcgatc tttggcaatg tgagggcccg 2700
 gaaacctggc cctgtttct tgacgagcat tccttaggggt ttccggccccc tgcggaaagg 2760
 aatgcaagggt ctgttgaatg tcgtgaagga agcagtttct ctggaaagctt cttgaagaca 2820
 aacaacgtct gttagcgcacc ttgcaggca gcggaacccc ccacccggcg acagggtgcct 2880
 ctgcggccaa aagccacgtg tataagatac acctgcaaaag ggggcacaaac cccagtgcca 2940
 cggtgtgagt tggatagtt tgaaaagagt caaatggc tccctcaagcg tattcaacaa 3000
 ggggctgaag gatgccaga agtacccca ttgtatgggat tctgtatctgg ggcctcggtg 3060
 cacatgtttt acatgtgtt agtcgaggat aaaaaaaacgtt ctggggccccc cgaaccacgg 3120
 ggacgtgggtt ttcccttggaa aaacacgata ataccatggg cggccggcgcc agcgtgctga 3180
 gccggccgca gctggacaag tgggagaaga acaagctgaa gcacatcgat tggggccagcc 3240
 gcctgtgtggaa gaccagcgcg ggctggccggcc agacccggcag cgaggagctg cccagccctgt 3300
 accagcgcgtt cggcgcgtt gacaccaagg acaagtccaa gaagaaggcc cggcggccgc 3360
 aggtgagcca gaactacccc atcgtgcaga tcagggccccc caccctgtaaac gctggggatgaa 3420
 aggtgatccc catgttgcac gccctgagcg tggatggccgc caccaggccg 3480
 tggatggccgc cggcggccgc cccggccgc aacagcagcc 3540
 aggtgagccca gaactacccc atcgtgcaga acctgcaggcc cccggccgc 3600
 tggatggccgc cggcggccgc cccggccgc aacagcagcc 3660
 aggtgatccc catgttgcac gccctgagcg tggatggccgc cccggccgc 3720
 tggatggccgc cccggccgc aacagcagcc 3780
 aggtgatccc catgttgcac gccctgagcg tggatggccgc cccggccgc 3840
 tggatggccgc cccggccgc aacagcagcc 3900
 tggatggccgc cccggccgc aacagcagcc 3960
 tggatggccgc cccggccgc aacagcagcc 4020
 tggatggccgc cccggccgc aacagcagcc 4080
 tggatggccgc cccggccgc aacagcagcc 4140
 tggatggccgc cccggccgc aacagcagcc 4200
 tggatggccgc cccggccgc aacagcagcc 4260
 tggatggccgc cccggccgc aacagcagcc 4320
 tggatggccgc cccggccgc aacagcagcc 4380
 tggatggccgc cccggccgc aacagcagcc 4440
 tggatggccgc cccggccgc aacagcagcc 4500
 tggatggccgc cccggccgc aacagcagcc 4560
 tggatggccgc cccggccgc aacagcagcc 4620
 tggatggccgc cccggccgc aacagcagcc 4680
 aagtctaga 4689

<210> 75

<211> 4472

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp160.modUS4.delV1/V2.gag.modSF2

<400> 75

gaattcgcca ccatggatgc aatgaagaga gggctctgtgtgtgtggaa 60
gcagtcttcg tttcgccca cggcaccaccatcgtggatggccgttgc 120

cccggtgtgga aggaggcccac caccacccctg ttctgcgcca gcgacgccaa ggcttacaag 180
ggcaggcccc acaaactgtg ggcacccac gctctgcgtgc ccaaccgaccc caaccccccag 240
gaggtgaacc tgaccaacgt gacccagaac ttcaacatgt ggaagaacaaa catggtgag 300
cagatgcattt aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgggcgcc 360
ggccaggcct gcccccaaggt gagttcagag cccatccca tccactactg cgccccccgc 420
ggcttcgcca tcctgaagtg caaggacaag aagttcaacg gcacccggccc ctgcaagaac 480
gtgagcaccc tgcagtgcac ccacggcattt cggccctgtgg tgacccatcca gctgctgtg 540
aacggcagcc tggccgagga ggagatcgtt ctgegetcccg agaacttcac cgacaacgc 600
aagaccatca tcgtgcagct gaacgagtcc gtggagatca actgcattccg ccccaacaac 660
aacacgcgta agagcatccaa catggcccccc ggcgcgecct tctacgcccac egggcgcac 720
atcggcgaca tccgcccaggc caactgcacaa atcagcaagg ccaactggac caacaccctc 780
gagcagatcg tggagaagct ggcgcagcag ttccgcacaaaca acaagaccat catcttcaac 840
agcagcagcg gccggcgaccc cggatcgtt ttccacagct tcaactgcggg cggcgcagtt 900
ttctactgtca acaccagccaa gctgttcaac agcaccttggaa acatcaccggaa ggaggtgaac 960
aagaccaagg agaacgcac catcacttcc ccttgcgcac teccgcagat catcaacatg 1020
tggcaggagg tggcaaggccatgtacgc cccccccatcc gggcccgat ccaactgac 1080
agcaatatta cccggctgtt gtcggccggc aacatgaaagg acaactggcg cggcaccac 1140
gacaccgaga ccttccggccc cggcgccggc aacatgaaagg acaactggcg cggcaccac 1200
tacaagtaca aggtggtgcg catcgagcccc ttggccgtgg cccccccatcc gggccaggat 1260
cgcggtgtgc agegcgagaa ggcgcgcgtt ggcctggcg ccctgttcat cggccaccac 1320
ggcgcgcgcg ggagcaccat gggcgccgcg tccgtgaccc tgaccgtgca ggccgcgcac 1380
ctgctgagcg gcacgtgcac gcacgcacaa aacctgtgc ggcgcategaa ggcccgac 1440
cacctgtgc agctgaccgtt gtggggcattt aagcagctgc äggcccgcat cctggccgtg 1500
gagcgttacc tgaaggacca gcaactgtgtt ggcacatctggg gtcgcagcggg caagctgtatc 1560
tgcaccacca ccgtgcctt gAACAGCAGC tggagcaaca agagcctgtac cgagatctgg 1620
gacaacatga cctggatggaa gtggggagcgc gacatcgccaa actacaccggg cctgatctac 1680
aacctgtacg agatcgccca gaaaccagcag gagaagaacg ageaggagct gctggagctg 1740
gacaagtggg ccagcctgtt gaaactggttt gacatcacca actggctgtt gtagatccgc 1800
atcttcatca tgatcggtttt ccggctgtatc ggcctgcgcä tctgttgcg cgtgtgtgac 1860
atcgtaacc gctgtgcgcg gggctacagc cccatcagcc tgcagaccgg cctgccccgg 1920
cagcgcgcgc ccgacccggcc cgggggcattt gaggaggagg ggcgcgcac 1980
cgcacgaaacc gctctgtgtca cggcctgtgtt gcccgtatctt gggacgcacct ggcgcagctg 2040
tgcctgttca gctaccaccc cttgcgcgcac tctgtgtca tcgtggcccg catctgtggag 2100
ctgtggggcc gccggcgctt ggaggccctt aagactgtt ggaacctgt gcaactgtt 2160
agccaggagc tgaagagcag cccgtgaccc ttgttcaacg ccaccccat cggcgttccgt 2220
gagggcaccg accgcattatc cggatcgtt cagcgcattt tccgcgcgtt gatccacatc 2280
ccccgcgcga tccggccaggg cctggagcgc gcccgtt aagatatccg atctctttaga 2340
gaattccccc ccccccccccc cccccccctt ccttcccccc cccctaacgt tactggccga 2400
agccgcttgg aataaggccg gtgtgcgtttt gtctatatgtt tattttccac catattgcgg 2460
tcttttggca atgtggggc cccggaaacctt gcccctgtct tcttgacgag catttccttagg 2520
ggtctttccc ctctgcacaa aggaatgcac ggtctgttga atgtcggtaa ggaagcagtt 2580
ctctctggaa gtttttgcac acaaacaacg tctgtatcga cccttgcac gcaagcggaaac 2640
cccccaccc ggcacagggtt cctctgcggc caaaaggccac gtgtataaga tacacctgc 2700
aaggcggcac aaccccaagtcc acacgttgc agttggatag ttgtggaaag agtcaaatgg 2760
ctctcttccaa gctgtatttcaaa caaggggctt gaggatcccc agaagggtacc ccattgtatg 2820
ggatctgtatc tggggccttgc gtgcacatgc ttatcatgtt ttagtgcag gttaaaaaaa 2880
cgctcttagcc ccccgaaacca cggggacgtt gttttccctt gaaaaacacg ataataccat 2940
gggcgcgcgc gccagcgtgc tgagcggcgcc cgagctggac aagtgggaga agatccgcct 3000
gccccccgcg gcaagaagaaga agtacaagctt gaaagcacatc gtgtggggca gccgcgagct 3060
ggagcgttcc gccgttgcacc cccggctgtt gggagacccg gagggtctgc gccagatct 3120
ggggccagctt cggccaggcc tgcagacggc cagcggagg ctgcgcagecc tgcacac 3180
cggtggccacc ctgtactgcg tgcaccaggcc catcgacgtc aaggacacca aggaggccct 3240
ggagaagatc gaggaggagc agaacaagtc caagaagaag gcccagcagg cccgcgcgc 3300
cgccggcacc ggcaacagca gccaggttag ccagaactac cccatgtgc agaactgc 3360
ggggccagatg gtgcaccagg ccatcagccc cccgcacccctt aacgccttgg tgaagggtgt 3420
ggaggagaag gccttcagcc cccggatgtatc cccatgttca aggccttgc gcgaggggcgc 3480
caccctccatc gaccatca cgtatgttgc caccgtggc ggcaccacagg cccgcacatgc 3540
gatgtgttca gacccatca acggaggaggcc cggcgttgc gaccgcgtgc aacccgttgc 3600
cgccggcccc atcgccccccg gccagatgcg cggcccccgc ggcacgcaca tcgcggccac 3660

caccagcacc	ctgcaggaggc	agatcggtctg	gatgaccaac	aacccccc	tccccgtggg	3720
cgagatctac	aagegggtgg	tcatcctggg	cctgaacaag	atcgtgcgg	tgtacagccc	3780
caccagcatc	ctggacatcc	gccaggcccc	caaggagccc	ttcccgact	acgtggaccg	3840
cttctacaag	accctgcgcg	ctgagcgggc	cagccaggac	gtgaagaact	ggatgaccga	3900
gaccctgctg	gtgcagaacg	ccaaccccg	ctgcaagacc	atccctgaagg	ctctcgcccc	3960
cgccggccacc	ctggaggaga	tgatgaccgc	ctgcccaggc	gtgggcggcc	ccggccacaa	4020
ggcccgctgt	ctggccgagg	cgatgagcca	ggtgaccaac	ccggcgacc	tcatgtgca	4080
gcgcggcaac	ttccgcaccc	agcggaaagac	cgtcaagtgc	ttaactgcg	gcaaggaggg	4140
ccacaccgccc	aggaactgcc	gcgcggcccc	caagaaggc	tgtggcgct	gcggccgcga	4200
gggccaccag	atgaaggact	gcacccggc	ccaggccaa	ttccctggca	agatctggcc	4260
cagctacaag	ggccgcggcc	gcaacttcct	gcagagccgc	cccgagccca	ccggcccccc	4320
cgaggagagc	ttccgcgtcg	gcgaggagaa	gaccaccccc	agccagaagc	aggagcccat	4380
cgacaaggag	ctgtacccccc	tgaccagcc	gcmcagctg	ttcggcaacg	accccgacag	4440
ccagtaagaa	ttcagactcg	agcaagtcta	qa			4472

<210> 76

<211> 4608

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modSF162.delV2.gag.modSF2

<400> 76

gaattcggca	ccatggatgc	aatgaagaga	gggctctgct	gtgtgctgct	gctgtgtgga	60
gcagtcttcg	tttgcggccag	cgcgggtggag	aagctgtggg	tgaccgtgta	ctacggcggtg	120
cccggtgtgg	aggaggccac	caccacccctg	ttctgcggcca	gcgacgccaa	ggcctacgac	180
accgagggtgc	acaacgtgtg	ggccacccac	gcctgegtgc	ccaccggaccc	caaccccccag	240
gagatcggtgc	tggagaacgt	gaccgagaac	ttcaacatgt	ggaagaacaa	catggtggag	300
cagatgcacg	aggacatcat	cagcctgtgg	gaccagagcc	tgaagccctg	cgtgaagctg	360
acccccctgt	gcgtgaccct	gcactgcacc	aacctgaaga	acgccaccaa	caccaagagc	420
agcaacttgg	aggagatgga	cccgccggcag	atcaagaact	gcagcttcaa	ggtgggcgcc	480
ggcaagctga	tcaactgcaa	caccagcgtg	atcacccagg	cttgccccaa	ggtgagcttc	540
gagcccatcc	ccatccacta	ctgggggggg	ggcggttccg	ccatccctgaa	gtgcaacgac	600
aagaagtca	acggcagcgg	ccccctgcacc	aacgtgagca	ccgtgcagtg	caccacccgc	660
atccggcccg	tggtagccac	ccagctgtgc	ctgaaggccg	gcctggccg	ggggggcggtg	720
gtgatccgca	gggagaactt	caccgtacaa	gccaagacca	tcaactgtgca	gctgaaggag	780
agcgtggaga	tcaactgcac	ccggggggaa	aacaacaccc	gcaagagcat	caccatcgcc	840
cccgccgcg	ccttctacgc	caccggcgcac	atcatccgg	acatccgcac	ggcccactgc	900
aacatcagcg	gcgagaagtg	gaacaacacc	ctgaagcaga	tcgtgaccac	gctgcaggcc	960
cagttcggca	acaagaccat	cgtgttcaag	cagagcagec	ggggcgaccc	cgagatcggt	1020
atgcacagct	tcaactgggg	ccggcgagttc	ttctactgca	acacgacccca	gtgttcaac	1080
agcaccttgg	acaacaccat	ccggggggaa	aacaccaacg	gcacccatcac	cctggccctgc	1140
cgcatacaagc	agatcatcaa	ccgctggcag	gaggtgggca	aggccatgt	cgccccccccc	1200
atccgcggcc	agatccgtg	cagcagcaac	atcacccggcc	tgctgctgac	ccgcgacggc	1260
ggcaaggaga	tcagcaacac	caccgagatc	ttccggggcc	ggggcgggcga	catgcgcgac	1320
aactggcgca	gcgagctgta	caagtacaag	gtggtaaga	tcgagccccct	gggcgtggcc	1380
cccaccaagg	ccaaagcggcg	cgtggtgca	cgcgagaagc	gcccgggtgac	cctggggcgcc	1440
atgttcttgg	gtttcttggg	ccggccggcgc	agcaccatgg	gcccggcag	cctgaccctg	1500
accgtgcagg	cccgccagct	gtcgagcggc	atcggtcagc	agcagaacaa	cctgctgcgc	1560
gccccatcgagg	cccagcagca	ctgtgtcag	ctgaccgtgt	ggggcatcaa	gcagctgcag	1620
gtccggcggtc	ttggccgtgg	ggcgatccctg	aaggaccagc	agctgctggg	catctggggc	1680
tgccatcgccgca	agctgtatctg	caccacggcc	gtggccctgg	acgccagctg	gagaacaacag	1740
agccctggacc	agatctggaa	caacatgacc	tggatggagt	gggagcgcga	gatgcacaac	1800
tacaccaacc	tgatctacac	cctgtatcgag	gagagccaga	accaggcagg	gaagaacagag	1860
caggagctgc	tggagctgga	caagtggggc	agcctgtgg	acttggttcga	catcagcaag	1920
tggctgttgtt	acatcaagat	cttcatcatg	atcggtggcg	gcctgggtggg	cctgctgcatc	1980
gtgttccacccg	tgctgagcat	ctgtgaaccgc	gtgcggccagg	gctacagccc	cctgagcttc	2040

cagaccggcgt tccccggccc ccggggggcc gacggccccg agggcatcgaa ggaggaggggc 2100
ggcgagcgcg accgggaccc cagcagcccc ctggtgcacg gcctgtggc cctgatctgg 2160
gacgacctgc gcagggctgtg cctgttcage taccacccgc tgccgaccc gatccctgatc 2220
gecgccccca tcgtggagct gctggggcgc cgccggctggg aggccctgaa gtactggggc 2280
aacctgtgc agtactggat ccaggagctg aagaacagcg ccgtgacect ttctgacgccc 2340
atcgccatcg ccgtggccga gggcacccac cgcatcatcg aggtggccca gecatcgccc 2400
cgcccttcc tgacatccc cccggccatc cgccagggtc tcgagcgcgc cctgctgtaa 2460
ctcgagcaag tctagagaat tccggggggggg ccccccccccc ccctctcccc ccccccccccc 2520
taacgttact ggccgaagcc gcttggaaata aggccggtgt gcgtttgtct atatgttatt 2580
ttccaccata ttggcgttctt ttggcaatgt gagggccccc aaacctggcc ctgttcttctt 2640
gacgagcatt cctaggggtc tttccccctct cggccaaagga atgcaaggtc tggtgaatgt 2700
cgtgaaggaa gcaggccctc tggaaagcttc ttgaagacaa acaaactgtcg tagcaccct 2760
ttgcaggcag cggaaacccccc cacctggcga caggtgeetc tgccggccaaa agccacgtgt 2820
ataagataca cctgc当地 cggcacaacc ccagtgcac gttgtgagtt ggatagttgt 2880
ggaaagagtc aaatggctct cctcaagcgtt atccaacaag gggctaagg atgcccagaa 2940
ggtaaaaaat ttgtatggat ctgatctggg gcctcggtc acatgttta catgtgttta 3000
gtcgagggtt aaaaaaacgtc tagggggggggg gaaccacccggg gacgtgggtt tcttttgaaa 3060
aacacgataa taccatgggc gcccggccca gctgtgcggag cggccggcagag ctggacaagt 3120
gggagaagat ccgc当地 cccggggccca agaaagaagta caagctgaag caccatcggt 3180
gggcccagccg cgagctggag cgcttgcggc tgaacccccc cctgctggag accagcgagg 3240
gctggccca gatctgggc cagctgc当地 ccagectgca gaccggcagc gaggagctgc 3300
gcagcctgta caacaccgtc gccaccctgt actgcgtgca ccagcgcaccc gacgtcaagg 3360
acaccaagga ggcctggag aagatcgagg aggagcagaa caagtccaaag aagaaggccc 3420
agcaggccgc cggccccc当地 ggcacccggca acagcagccca ggtgagccag aactacccca 3480
tcgtgcagaa cctgcaggcc cagatggtgc accagggccat cagccccccgc accctgaacg 3540
cctgggtgaa ggtgtggag gagaaggccct tcaagccccca ggtgatcccc atgtttagcg 3600
ccctgagcga gggccccc当地 ccccaggacc tgaacacgat gttgaacacc gttggccggcc 3660
accaggccgc catgcagatg ctgaaggaga ccatcaacga ggaggccccc gagtgggacc 3720
gctgtcaccc cgtgcacccccc ggccccatcg ccccccggccca gatgcgc当地 ccccgccggca 3780
gcfgacatcg cggcaccacc cggccaccctgc aggagcagat cggctggatg accaacaacc 3840
cccccatccc cgtggccgag atctacaagg ggtggatcat cctggggctcg aacaagatcg 3900
tgccggatgt cagccccacc agcatectgg acateccggca gggcccccaag gagcccttcc 3960
gcfgactacgt ggaccggctt tacaagaccc tgcgc当地 gcaaggccgc当地 cggacgtga 4020
agaaggatcg gaccggagcc ctgtgtgtc agaacggccaa ccccgactgc aagaccatcc 4080
tgaaggctct cggccccc当地 gccacccctgg aggagatgt gaccggccctgc caggccgtgg 4140
gcccggccccc ccacaaaggcc cgcgtgtgg cccggggccat gaggccaggccgg 4200
cgaccatcat gatgcagcgc ggcaacttcc gcaaccaggcg gaagaccgtc aagtgttcc 4260
actgcggcaa ggaggggccac accggccaggaa actggccggc ccccccggcaag aaggccgtgt 4320
ggcgctgcgg ccgc当地 caccagatga aggactgcac cgagcgc当地 gccaacttcc 4380
tgggcaagat ctggccccc当地 tacaaggggcc gccccggccaa cttcctgcaag agccggccccc 4440
agccccccccc cccccccccc当地 gagagcttcc gttcgccga ggagaagacc accccccagcc 4500
agaaggccggaa gccc当地 aaggagctgt accccctgac cagccctgc当地 agcctgttcc 4560
gcaacccccc cagcagccag taagaattca qactcgagca agtctaga 4608

<210> 77
<211> 1680
<212> DNA
<213> Human

<400> 77

```

ccccattagtc ctattggaaaac tgttaccaggta aaattaaagc caggaatggaa tggcccaaaaa 60
gtttagcaat ggccatttgac agaagaaaaaa ataaaaagcat tagtagagat atgtacagaaa 120
atggaaaagg aaggggaaaat ttccaaaaatt gggctgaaa atccatacaa tactccagta 180
tttgctataa agaaaaaaaga cagtactaaa tggagaaaaac tagtagatTT cagagaacct 240
aataaaaagaa ctcaagactt ctggaaagt cagttaggaa taccacacccc cgccagggtta 300
aaaaagaaaaa aatcgtaaac agtatggat gtgggtgtatg catacttttc agttccctta 360
gataaaagact tttagaaaagta tactgcattt accataaccta gtataaacaa tgagacaccca 420
gggatttagat atcagtacaaat tttgtctgcaca cagggatggaa aaggatcacc agcaatattc 480
caaagtagca tgacaaaaat ctttagaccc ttttagaaaaac agaaatccaga catagttatc 540

```

tatcaataaca tggatgattt gtatgttagga tctgacttag aaatagggca gcatagaac 600
 aaaatagagg aactgagaca gcacatctgttg aggtggggat ttaccacacc agacaaaaaa 660
 catcagaaaag aacctccatt cctttggatg ggttatgaac tccatctga taaatggaca 720
 gtacagccata taatgtcgcc agaaaaagac agctggactg tcaatgacat acagaagtta 780
 gtggggaaaat tgaattgggc aagtcaagatt tatgcaggga ttaaagtaaa gcagttatgt 840
 aaacteecttta gaggaaaccaa agactaaca gaagtaatac cactaacaga agaagcagag 900
 ctagaactgg cagaaaaacag ggagattcta aaagaaccag tacatgaagt atattatgac 960
 ccataaaaaag acttagtagc agaaatacacg aagcaggggc aaggccaatg gacatatcaa 1020
 atttatcaag agccattaa aaatctgaaa acagggaaatg atgcaaggat gagggggtgcc 1080
 cacactaatg atgtaaaaaca gttaacacag gcagtgcaaa aagtatccac agaaagcata 1140
 gtaatatggg gaaagattcc taaaattttaa ctacccatc aaaaaggaaac atgggaagca 1200
 tgggtggatgg agtattggca agtacaccttgg attcctgagt gggagttgt caataccccct 1260
 cccttagtga aattatggta ccagtttagag aaagaaccca tagtaggagc agaaactttc 1320
 tatgttagatg gggcagctaa tagggagact aaatttaggaa aagcaggata tttactgac 1380
 agaggaagac aaaaagttgt ctccatagct gacacaacaa atcagaagac tgaattacaa 1440
 gcaattcattc tagcttgca ggattcggga ttagaagtaa acatagtaac agactcacaa 1500
 tatgcatttag gaatcattca agcacaacca gataagatg aatcagagtt agtcagtcaa 1560
 ataataagac agttaataaaa aaaggaaaag gtctacctgg catgggtacc agcacacaaa 1620
 ggaattggag gaaatgaaca agtagataaa ttagtcagtg ctggaatcag gaaagtacta 1680

<210> 78

<211> 1865

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: GP1

<400> 78

gtcgacgcca ccatgggcgc cggcgccagc gtgctgagcg gcccggagct ggacaagtgg 60
 gagaagatcc gcctgcgccc cggcgcaag aagaagtaca agctgaagca catctgtgtgg 120
 gccagccgcg agctggagcg ctteccgtg aaccccgccc tgctggagac cagcgagggc 180
 tgccgcccaga tcctggccca gtcgcageccc agcctgcaga cccggcagcga ggagctgcgc 240
 agcctgtaca acaccgtggc caccctgtac tgctgcacc accgcacatcga cgtcaaggac 300
 accaaggagg ccctggagaa gatcgaggag gagcagaaca agtcaagaa gaaggcccag 360
 caggccgccc cggccgcccgg caccggcaac agcagccagg tgagccagaa ctacccatc 420
 gtgcagaacc tgcagggcca gatgggtgcac caggccatca gccccgcac cctgaacgccc 480
 tgggtgaagg tgggtggagga gaaggccttc agcccccagg tgatccccat gtcagcgcc 540
 ctgagcgagg gcccaccccc ccaggacctg aacacgatgt tgaacaccgt gggcgccac 600
 caggccgcca tgcagatgtc gaaggagacc atcaacgagg aggcccggca gtgggaccgc 660
 gtgcacccccc tgcacggccgg ccccatgcgc cccggccaga tgcgcgagcc cccggccagc 720
 gacatcgccg gcaccacccag caccctgcag gagcagatcg gctggatgac caacaacccc 780
 cccatccccg tggcggatg ctacaagcgcc tggatcatcc tgggcctgaa caagatcg 840
 cggatgtaca gccccccagg cttctggac atccggcagg gcccccaagga gcccctccgc 900
 gactacgtgg accgccttca caagacccctg cgcgtgagc aggcccggca ggacgtgaag 960
 aactggatga cggagacccct gctgggtgcag aacgcacccaa cccactgcac gaccatccctg 1020
 aaggctctcg gccccggcgc caccctggag gagatgtatca cccgcctgcca gggcgtgggc 1080
 gccccggcgc acaaggccccg cgtgctggcc gaggcgatgc gccaggtgac gaacccggcg 1140
 accatcatga tgcagcgccg caacttccgc aaccagcgca agaccgtcaa gtgcttcaac 1200
 tgccgcaagg agggccacac cgcaggaaac tgccgccc cccgcagaagaa gggctgctgg 1260
 cgctgcggcc gcgaggaca ccaaataaaa gattgcactg agagacaggc taatttttta 1320
 gggaaagatct ggccttccta caagggaaagg ccagggaaatt ttcttcagag cagaccagag 1380
 ccaacagccc caccagaaga gagttcagg tttggggagg agaaaaacaac tccctctcag 1440
 aagcaggagc cgatagacaa ggaactgtat cctttaactt ccctcagatc actctttggc 1500
 aacgacccct cgtcacagta aggatcgccg gccagctcaa ggaggcgctg ctgcacaccg 1560
 gcgccgacga caccgtgctg gaggagatga acctgcggcc caagtggaaag cccaagatga 1620
 tcggcgggat cgggggcttc atcaaggtgc ggcagtgacca ccagatcccc gtggagatct 1680
 gccggccacaa ggcctcgcc accgtgctgg tggggccac ccccgtaac atcatcgcc 1740
 gcaacctgct gacccagatc ggctgcaccc tgaacttccc catcagcccc atcgagacgg 1800

tggccgtgaa gctgaagccg gggatggacg gcccccaaggt caagcagtgg cccctgttaag 1860
aattc 1865

<210> 79
<211> 1865
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: GP2

<400> 79
gtcgacgcca ccatgggcgc ccgcgccagc gtgctgagcg gggcgagct ggacaagtgg 60
gagaagatcc geetgcccggcc cgccggcaag aagaagtaca agctgaagca catcggtgg 120
gccagccgcg agctggagcg ctgcggcgta aaccccgccg tgctggagac cagcgaggc 180
tgcggccaga tccctgggcca gtcgcagccc agcctgcaga cccgcagcga ggagctgcgc 240
agcctgtaca acacccgtggc caccctgtac tgctgcacc agcgcacgtca cgtcaaggac 300
accaaggagg ccctggagaa gatcgaggag gacgagaaca agtccaaagaa gaaggcccag 360
caggccgcgc cccgcgcggc caccggcaac agcagccagg tgagccagaa ctaccccatc 420
gtgcagaacc tgcaggggca gatgggtgcac caggccatca gccccccgcac cctgaacgccc 480
tgggtgaagg tggtgagga gaaggccctc agcccccggg tgatccccat gttcagcgcc 540
ctgagcgagg ggcgcacccccc ccaggacactg aacacgtatg tgaacacccgt gggggccac 600
caggccgcgc tgcagatgtc gaaggagacc atcaacgagg aggccgcggc gtgggacccgc 660
gtgcacccccc tgcacgcggc cccatcgcc gacatcgccg gcaccaccag caccctgcag 720
gacatcgccg gacatcgccg caccctgcag gacgagatcg gctggatgac caacaaccccc 780
cccatccccc tggcgagat ctacaagcgg tggatcatec tgggcctgaa caagatcg 840
cggtatgtaca gccccaccag catctggac atccgcagg gcccaagaga gcccctccgc 900
gactacgtgg accgcctcta caagaccctg cgcgctgagc aggccagcga ggacgtgaag 960
aactggatga ccgagaccct gctggtgcag aacgccaacc cgcactgcaaa gaccatccctg 1020
aaggctctcg gccccccgcg caccctggag gagatgtatc cgcctgcac gggcgtggc 1080
ggccccccgcg acaaggcccg cgtgtgtggcc gaggcgatga gccaggtgac gaacccggcg 1140
accatcatgtc tgcagcgccgg caacttcccg aaccagcgga agacccgtcaa gtgcttcaac 1200
tgcggcaagg agggccacac cgcaggaaac tgcggccccc cccgcacagaa gggctgtgg 1260
cgctgcggcc gcgaaaggaca cccaaatggaa gattgcactg agagacaggg taattttta 1320
ggaaagatctt ggccttcata caagggagg ccaggaaat ttcttcagag cagaccagag 1380
ccaaacagcccc caccagaaga gagttcaggg ttggggagg agaaaacaac tccctctcag 1440
aagcaggaggc cgatagacaa ggaactgtat cctttaactt ccctcagatc actttttggc 1500
aacgaccctt cgtcacagta aggtatcgaaaa ggcacactcaa ggaagcgctg ctcgatacag 1560
gagcagatgtc tacagtatta gaagaaatgtc atttgccagg aaaatggaaa caaaaaatgtc 1620
tagggggat cgggggcttc atcaagggtgc ggcagttacca ccagatacct gtagaaatct 1680
gtggacataaa agctataaggt acagtatttag taggacctac acctgtcaac ataattggaa 1740
gaaatctgtt gacccagatc ggctgcaccc tgaacttccc catcagccctt attgagacgg 1800
tgccctgtgaa gttgaagccg gggatggacg gcccccaaggt caagcaatgg ccattgttaag 1860
aattc 1865

<210> 80
<211> 2305
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
FS(+) .proinact.RTopt.YM

<400> 80
gcggccgcga aggacaccaa atgaaagatt gcactgagag acaggctaat ttttttaggg 60
agatctggcc ttccataaag ggaaggccag ggaattttct tcagagcaga ccagagccaa 120
cagccccacc agaagagacg ttccaggatgg gggaggagaa aacaactccc tctcagaacg 180
aggagccgat agacaaggaa ctgtatccctt taacttccctt cagatcactc tttggcaacg 240

accccctcgta acaataagga tcggggggca actcaaggaa gcgctgctcg atacaggagc 300
 agatgataca gtattagaag aaatgaattt gccaggaaaa tggaaaccaa aaatgatagg 360
 ggggatcggt ggcttcatca aggtgaggca gtacgaccag atacctgttag aaatctgtgg 420
 acataaaagct ataggtacag tattatggg acctcacact gtcaacataa ttgaaagaaa 480
 tctgttgacc cagatcggt gcaccttgaa cttccccatc agcccttattg agacggtgcc 540
 cgtgaagttt aagccggggta tggacggccc caaggtcaag caatggccat tgaccgagga 600
 gaagatcaag gcccctgggtt agatctgcac cgagatggag aaggagggca agatcagcaa 660
 gatcgcccccc gagaacccct acaacacccc cgtgttcgca atcaagaaga aggacagcac 720
 caagtggcgc aagctgggtt acttccgcga gctgaacaag cgacccaggc acttctggga 780
 ggtgcagctg ggcatcccccc accccccggg cttggacaag aagaagagcg tgaccgtct 840
 ggacgtgggc gacgcctact tcagegtgcc cttggacaag gacttccgca agtacaccgc 900
 cttcaccatc cccagcatca acaacgagac cccccggcatc cgctaccagt acaacgtct 960
 gccccagggc tggaaaggga gccccggcat cttccagagc agcatgacca agatcctgga 1020
 gccccccgc aaggcagaacc cccgacategt gatctaccag gccccctgt acgtggcag 1080
 cgacctggag atcggccacgc accgcaccaa gatcgaggag ctgeggcage acctgtctgc 1140
 ctggggcttc accacccccc acaagaagca ccagaaggag ccccccttcc tttggatggg 1200
 ctacgagctg caccggcaca agtggaccgt gcaagccatc atgctggccg agaaggacag 1260
 ctggaccgtg aacgacatcc agaagctggt gggcaagctg aactggccca gccagatcta 1320
 cggccggcattc aaggtgaagc agctgtgcaa gctgctgcgc ggcaccaagg ccctgaccga 1380
 ggtgatccccctt ctgaccggagg aggccgagct ggagctggcc gagaaccgcg agatcctgaa 1440
 ggagcccggtg caccgggtgt actacgaccc cagcaaggac ctggggccg agatccagaa 1500
 gcaggggccag gggcagtggc cttaccagat ctaccaggag ccccttcaaga acctgaagac 1560
 cggcaagttac gcccgcattc gggcgccca caccacgcgt gtgaagcgc tgaccgagc 1620
 cgtgcagaaag gtgagcaccgc agacgcattt gatctggggc aagatccccca agttcaagct 1680
 gccccatccaa aaggagacccctt gggagggctt gtggatggg tactgtcagg ccacctggat 1740
 ccccgagtgg gagttcgtga acacccccc cctggtaag ctgtgttacc agctggagaa 1800
 ggagcccatc gtggggcccg agaccttcta cttggacccgc gccgcaccaacc gcgagaccaa 1860
 gctgggcaag gccggctacg tgaccgaccg gggccggcag aaggtggta gcatgcggca 1920
 caccaccaac cagaagaccc agtgcaggc catccacccgc gccctgcagg acagggcct 1980
 ggaggtgaac atcgtgaccg acagccgatc cggccatc atcatccagg cccagcccgaa 2040
 caagagcgag agcgagctgg tgagccagat catcgaggcgt ctgatcaaga aggagaaggt 2100
 gtacctggcc tgggtggccccc cccacaagggtt catggccgc aacgaggcagg tgacaagct 2160
 ggtgagcgcc ggcattccca aggtgtt cctgaacggc atcgatggcg gcatcgatgtat 2220
 ctaccagtac atggacgacc ttttttttttggggcc cctaggatcg attaaaagct 2280
 tccccgggtt agcaccgggtt aatttccatc 2305

<210> 81

<211> 229

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
FS(+).proinact.RTopt.YMWM

<400> 81

gggccgcgca aggacaccaa atgaaagatt gcactgagag acaggctaat ttttttaggga 60
 agatctggcc ttcctacaag ggaaggccag ggaatttttct tcagagcaga ccagagccaa 120
 cagccccacc agaagagagc ttccagggtt gggaggagaa aacaactccc tctcagaagc 180
 aggagccgtt agacaaaggaa ctgtatccct taacttccct cagatcactc tttggcaacg 240
 accccctcgta acaataagga tcggggggca actcaaggaa gcgctgctcg atacaggagc 300
 agatgataca gtattagaag aaatgaattt gccaggaaaa tggaaaccaa aaatgatagg 360
 ggggatcggt ggcttcatca aggtgaggca gtacgaccag atacctgttag aaatctgtgg 420
 acataaaagct ataggtacag tattatggg acctcacact gtcaacataa ttgaaagaaa 480
 tctgttgacc cagatcggt gcaccttgaa cttccccatc agcccttattg agacggtgcc 540
 cgtgaagttt aagccggggta tggacggccc caaggtcaag caatggccat tgaccgagga 600
 gaagatcaag gcccctgggtt agatctgcac cgagatggag aaggagggca agatcagcaa 660
 gatcgcccccc gagaacccct acaacacccc cgtgttcgca atcaagaaga aggacagcac 720
 caagtggcgc aagctgggtt acttccgcga gctgaacaag cgacccaggc acttctggga 780

ggtgcagctg ggcatcccc accccggcgg cctgaagaag aagaagagcg tgaccgtgct 840
 ggacgtgggc gacgcctact tcagcgtgcc cttggacaag gacttccgca agtacaccgc 900
 cttcaccatc cccagcatca acaacgagac cccggcata cgctaccagt acaacgtgct 960
 gccccagggc tggaaaggca gccccggcat cttccagagc agcatgacca agatcctgga 1020
 gcccctccgc aagcagaacc ccgacatcgat gatctaccag gccccctgt acgtggcag 1080
 cgacctggag atcggccacg accgcaccaa gatcgaggag ctgcggcage acctgctgct 1140
 ctggggcttc accacccccc acaagaagca ccagaaggag ccccccattcc tggccatcga 1200
 gtcgcacccc gacaagtggc cctgtcagcc catcatgtc cccgagaagg acagctggac 1260
 cgtgaacgc atccagaagc tgggtggcaaa gctgaacttg gccagccaga tctacgccc 1320
 catcaagggt aagcagctgt gcaagctgtc ggcggcacc aaggccctga ccgaggtgtat 1380
 cccctggacc gaggggccg agctggagct ggccgagaac cgcgagatcc tgaaggagcc 1440
 cgtgcacggc gtgtactacg accccagcaa ggacctgtt gccgagatcc aagaaggcagg 1500
 ccaggcccg tggacctacc agatctacca ggagcccttc aagaaccttga agaccggcaa 1560
 gtacgcccgc atgegcggcg cccacaccaa cgacgtgaag cagctgaccg aggcgtgca 1620
 gaaggtgagc accgagagca tcgtgatctg gggcaagatc cccaaatcca agtgcggccat 1680
 ccagaaggag acctgggagg cctgggtggat ggagtaactgg caggccacctt ggtatcccc 1740
 gtgggagttc gtgaacaccc ccccccctgtt gaaagctgtgg taccagctgg aagaaggagcc 1800
 catcggtggc gccgagaccc tctacgttga cggccggcc aaccggaga ccaagctggg 1860
 caaggccggc tacgtgaccg accggggccg gcagaagggt gtgagcatcg ccgacaccac 1920
 caaccagaag accgagctgc aggccatcca cctggccctg caggacagcg gcctggaggt 1980
 gaacatcgta accgacagcc agtacgcctt gggcatcatc caggcccagc ccgacaagag 2040
 cgagagcggc ctggtgagcc agatcatcg gcaagctgate aagaaggaga aggtgtaccc 2100
 ggcctgggtg cccggccaca agggcatcg ccggcaacggc caggtggaca agctggtgag 2160
 cgccggcattc cgcaagggtc tggctctgaa cggcatcgat ggccgcatcg tcatctacca 2220
 gtacatggac gacctgtacg tggcagcgg cggccctagg atcgattaaa agcttcccg 2280
 ggcttagcacc ggtgaattc 2299

<210> 82

<211> 2306

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
FS(-).protmod.RTopt.YM

<400> 82

gccggccgcga aggacaccaa atgaaagatt gcaactgagag acaggctaat ttcttccgcg 60
 aggacctggc cttccctgcag ggcacggccc gcgagttcag cagcgagcag acccgccca 120
 acagccccac cccggccgcag ctgcagggtgt gggggccgcga gaacaacgc ctgagcgagg 180
 cccggccgcga cccggccaggc accgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
 gccccctgtt gaccatcagg atccggccgc agctcaaggaa ggcgctgctc gacaccggcg 300
 cccgacgacac cgtgctggag gagatgaacc tggccggccaa gtggaaagccc aagatgatcg 360
 gcccggatcg gggcttcatc aaggtgcggc agtacgacca gatccccgtg gagatctgct 420
 gcccacaaggc catggcacc gtgttgtgg gccccacccc cgtgaacatc atccggccgc 480
 acctgctgac ccagatcgcc tgccacccat acttcccccatt cagcccccattc gagacgggtgc 540
 cccgtgaagctt gaaaggccggg atggacggcc ccaagggtcaa gcagtggccc ctgaccggagg 600
 agaaagatcaa gggcccttttgg gagatctgca ccgagatgga gaaggaggccg aagatcagca 660
 agatcgccccc cgagaaccccc tacaacaccc cccgtgttgc catcaagaag aaggacagca 720
 ccaagtggcg caagtttgtt gacttccgcg agcttgcacaa ggcgcaccccg gacttctggg 780
 aggtgcagctt gggccatcccc caccggccg cccgtgaagaa gaagaagagc gtgaccgtgc 840
 tggacgtggg cccgttgc ttcagcgtgc ccctggacaa ggacttccgc aagtacacccg 900
 ctttcaccat ccccgccatc aacaacgaga ccccccggcat ccgttaccatc tacaacgtgc 960
 tggcccccaggc ctggaaaggc accccggcata tcttcccgag cagcatgacc aagatcctgg 1020
 agcccttccg caagcagaac cccgacatcg tcatctacca ggccccctgt tacgtggca 1080
 ggcacccgtt gatcgcccg caccgcacca agatcgagga gctgcggccag caccctgtgc 1140
 gctggggctt caccacccccc gacaagaagc accagaaggc gcccccttc ctgtggatgg 1200
 gctacgagctt gcacccggac aagtggaccg tggccatcgatc tcatgtgcggc gagaaggaca 1260
 gctggaccgtt gacgcacatc cagaagctgg tggcagccat cagatgttcc gagaaggatct 1320

acggccggcat caagggtgaag cagctgtgca agctgctgcg cgccaccaag gccctgaccg 1380
 aggtgatccc cctgaccgag gaggccgagc tggagctggc cgagaaccgc gagatcctga 1440
 aggagcccggt gcacgagggt tactacgacc ccagcaagga cctggggcc gagatccaga 1500
 agcagggcca gggccatgg acttaccagaa tctaccaggaa gccctcaag aacctgaaga 1560
 cccggcaagta cggccatgc cgccggcccc acaccaacga cgtgaagcag ctgaccgagg 1620
 cccgtgcagaa ggtgagcacc gagactcgt tgatctgggg caagatcccc aagttcaagc 1680
 tggccatcca gaaggagacc tggaggcct ggtggatggta gtactggcag gccacctgga 1740
 tccccggatgg ggagtctgt aacacccccc ccctggtaa gctgttgtac cagctggaga 1800
 aggagcccat cgtggccgc gagacattct acgtggacgg cgccgcaac cgccgagacca 1860
 agctgggcaa ggccggctac gtgaccgacc gggggccggca gaaggtggtg agcatcgecg 1920
 acaccaccaa ccagaagacc gagactgcagg ccatccaccc ggcctgcag gacagccggc 1980
 tggaggtgaa catcgtgacc gacagccagt acggccctggg catcatccag gcccagcccg 2040
 acaagagcga gagcggatgt gtgagccaga tcatcgagca gctgatcaag aaggagaagg 2100
 tgtacctggc ctgggtggcc gccccacaagg gcatcgccgg caacgagcag ttggacaagc 2160
 tggtgagcgc cggcatccgc aaggtgtgt tctgaacgg catcgatggc ggcatacgta 2220
 tctaccagta catggacgac ctgtacgtgg gcagccggg ccctaggata gattaaaagc 2280
 tccccggggc tagcaccggtaa attc 2306

<210> 83

<211> 2300

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

PS(-).protmod.RTopt.YMWM

<400> 83

gcgccgcgca aggacaccaa atgaaagatt gcaactgagag acaggctaatttcccg 60
 aggacctggc cttectgcag ggcaaggccc gcgagttcag cagcgacgac acccgccca 120
 acagccccac ccggccgcag ctgcagggtgt gggggccggca gaacaacagc ctgagcgagg 180
 cccggccgcgca ccggccaggcc accgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
 gccccctggt gaccatcagg atccggggcc agctcaagga ggcgtgtctc gacacccggc 300
 cccgacgacac cgtgtggag gagatgaacc tgcccgccaa gtggaaagccc aagatgatcg 360
 gcccggatcg gggcttcatc aaggtgcggc agtacgacca gatccccgtg gagatctgcg 420
 gccacaaggc catcgccacc gtgtgtgtgg gccccacccc cgtgaacatc atccggccgc 480
 acctgtgtac ccagatccgc tgcaaccctga acttcccccatt cagccccatc gagacgggtgc 540
 ccgtgaagct gaagccgggg atggacggcc ccaaggtaa gcagtggccc ctgaccgagg 600
 agaagatcaa ggccttggt gagatctgcg cggagatgga gaaggaggccc aagatcagca 660
 agatcggccc cgagaaccccc tacaacaccc cctgtgttcgc catcaagaag aaggacacga 720
 ccaagtggcg caagctggtg gacttccgc agtctgaacaa gcgccacccatc gacttctgg 780
 aggtgcagct gggcatcccc cacccccggcgc gcttgaagaa gaagaagagc gtgaccgtgc 840
 tggacgtggc cgacgcctac ttctacgtgc ccctggacaa ggacttccgc aagtacacccg 900
 ctttccacccat ccccgacatc aacaacgaga ccccccggcat cctgttccacccatc tacaacgtgc 960
 tgccccaggg ctggaaaggc agcccccggca tcttccacccat cagcatgacc aagatctgg 1020
 agcccttccg caagcagaac cccgacatc tgatctacca ggcccccttc tacgtgggca 1080
 gcgacctggc gatcgccag caccgcacca agatcgagga gtcgcggcag cacctgtgc 1140
 gctggggctt caccaccccc gacaagaagc accagaaggaa gcccccccttc ctgccccatcg 1200
 agctgcaccc cgacaagtgg accgtgcagc ccacatcatgt gccccgagaag gacagctgg 1260
 ccgtgaacga catccagaag ctggggccgca agtgcactg ggcacggccag atctacgcgc 1320
 gcatcaaggtaa gaagcagctg tgcaagctgc tgccggccac caaggccctg accgaggtga 1380
 tccccctgac cgaggaggcc gagctggagc tggccggagaa cccgcgagatc ctgaaaggagc 1440
 cccgtgcacgc ggtgtactac gaccccgacca aggacctggt ggcacggccatc cagaagcagg 1500
 gccaggccgcgtt gacccatc cagatctacc aggagccctt caagaacccatc aagacccggc 1560
 agtacggccgcg catcgccggc gccccacacca acgacgtgaa gcagctgacc gaggccgtgc 1620
 agaaggttag caccggagac atcgatctt gggggcaagat ccccaagttc aagctgcccc 1680
 tccagaagga gacctgggg gcttggatggta tggagttactg gcaggccacc tggatcccc 1740
 agtggggatgtt cgtgaacacc ccccccctgg tgaagctgtg gtaccagctg gagaaggagc 1800
 cccatcgatggc cggccgagacc ttctacgtgg acggccggccgc caaccgcgag accaagctgg 1860

gcaaggccgg	ctacgtgacc	gaccggggcc	ggcagaaggt	ggtaggcattc	ggcgacacca	1920
ccaaccagaa	gaccggagctg	caggccatcc	acctggccct	gcaggacagc	ggcctggagg	1980
tgaacatctgt	gaccggacagc	cagtacgccc	tgggcatcat	ccaggccccag	cccgacaaga	2040
gcgagagcga	gctggtgagc	cagatcatcg	agcagctgtat	caagaaggag	aagggtgttacc	2100
tggcctgggt	gcccggccac	aagggcatcg	ggggcaacga	gcagggtggac	aagctggtga	2160
gcgcggccat	ccgcaaggtg	ctgttcctga	acggcatcgat	tggcggcatc	gtgtatctacc	2220
agtacatgg	cgactctgtac	gtgggcagcg	ggggccctag	gatcgattaa	aagcttcccc	2280
qqqctaqaac	cqqtqaattc					2300

<210> 84
<211> 2312

<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
FS(-).protmod.RTopt(+)

<400> 84

gggcccgccgca aggacaccaa atgaaaaggatt gcactgagag acaggctaat ttcttcgcg 60
 aggaccttgc cttcctgcgag ggcaaggccc gcgagtttag cagcgagcag acccgccca 120
 acagccccac ccggccgcgag ctgcagggtgt gggggcggcga gaacaacagc ctgagcgagg 180
 cccggcgcgca ccggccagggc accgtgagct tcaacttccc ccagatcacc ctgtggcaggc 240
 gcccccttgtt gaccatcagg atcggcggcc agctcaagga ggcgctgctc gacaccggcg 300
 ccgacgacac cgtgtggag gagatgaacc tgcccgccaa gtggaaagccc aagatgatcg 360
 gggggatcgg gggcttcata aagggtgcggc agtacgacca gatccccgtg gagatctgcg 420
 gccacaaggc catcgccacc gtgtctggg gccccacccc cgtgaacatc atcgccgcga 480
 acctgtgtac ccagatcgcc tgacccctga acttccccat cagccccatc gagacggtg 540
 ccgtaagct gaagccgggg atggacggcc ccaaggtcaa gcagtggccc ctgaccgagg 600
 agaagatcaa ggcccttgtt gagatctgca ccgagatgga gaaggagggc aagatcagca 660
 agatcgcccc cgagaacccc tacaacaccc cctgttgcg catcaagaag aaggacacga 720
 ccaagtggcg caagttgtgt gacttccggc agtgaacaa ggcgcacccag gactttctgg 780
 aggtgcagct gggcatcccc cccccccggc gcttggaaaaaa gaagaagagc gtgaccgtgc 840
 tggacgtggg cgacgcttac ttcaagttgc ccctggacaa ggacttccgc aagtacaccg 900
 ccttcacat ccccaacatc aacaacgaga cccccggcat ccgttaccag tacaacgtgc 960
 tgccccaggc ctggaaaggc agccccggca tttttccagag cagcatgacc aagatctgg 1020
 agcccttcgg caagcagaac ccccgacatcg tgatctacca gtacatggac gacctgtacg 1080
 tgggcagega cctggagatc ggcacggcacc gcaccaagat cgaggagctg cgccagcacc 1140
 tgctgcgtcg gggcttcacc accccccggaca agaagcacca gaaggagccc cccttcctgt 1200
 ggatggggcta cgagctgcac cccgacaaatg ggaccgtgca gcccacatcg ctgcccggaga 1260
 aggacagctg gaccgtgaac gacatccaga agtgggtggg caagctgaac tggggccagcc 1320
 agatctacgc cggcatcaag gtgaaggcgc tttgtcaagct gcttgcggc accaaggccc 1380
 tgaccggaggt gatccccctg accggaggagg ccttggatcc gcttggccgag aaccggcggaga 1440
 tcctgaagga gcccgtgcac gaggtgtact acgaccccg caagacatcg gtggccggaga 1500
 tccagaagca gggccagggc cagtggaccc accagatcta ccaggagccc ttcaagaacc 1560
 tgaagacccgg caagtacgc cgcacatgcgcg gcccacac caacgcacgtg aagcagctga 1620
 cccggccgt gcagaagggtg accacccggaga gcatcgatgt ctggggcaag atccccaaatg 1680
 tcaactgtcc catccacaaag gagaccttggg aggccctgtgt gatggagttac tggcaggccca 1740
 ccttggatccc cgagtggggg ttcgttgcacca cccccccctt ggttggatcg tggtaccaggc 1800
 tggagaagga gcccacatcg ggcggccggaga cttttctacgt ggacccggcc gccaacccggc 1860
 agaccaaggtt gggcaaggcc ggttacgtga cccggccggg cccggcagaat gtgggtggac 1920
 tcggccgacac caccaacccggc aagacccggc tgcggccat ccacatggcc ctggcaggaca 1980
 gggcccttggg ggttggatccatc gtggccggaca gccagatcg ccttggccatc atccaggcccc 2040
 agcccgacaa gagcgagacg gagctgggtga gccagatcat cgagcagctg atcaagaagg 2100
 agaagggtgtt cctggcttgg gtggccggcc acaaggccat cggccggcaac gaggcagggtgg 2160
 acaagctgtt gggccggc acccgcaagg tttgttccat gaaacggccatc gatggccggca 2220
 tcgtgtatcta ccagtatcg gacgacatgt acgttggccatc cggccggccctt aggatcgatt 2280
 aaaagcttcc cggggcttgc accgggtgaat tc 2312

<210> 85

<211> 306

<212> DNA

<213> Human immunodeficiency virus

<400> 85

atggagccag tagatcctag attagagccc tggaaagcatt caggaagtca gcctaagact 60
 gcttgcataaa attgtctattg taaaaagggtgt tgctttcatt gccaagtttg ttcataaca 120
 aaaggcttag gcatctccta tggcaggaag aagcggagac agcgcacgaag agcttcctcca 180
 gacagtggagg ttcatcaagt ttctctacca aagcAACCG cttcccagcc ccaaggggac 240
 ccgcacaggcc cgaagaatac gaagaagaag gtggagagag agacagagac agatccagtc 300
 cattag

306

<210> 86

<211> 101

<212> PRT

<213> Human immunodeficiency virus

<400> 86

Met	Glu	Pro	Val	Asp	Pro	Arg	Leu	Glu	Pro	Trp	Lys	His	Pro	Gly	Ser
1															

5

10

15

Gln	Pro	Lys	Thr	Ala	Cys	Thr	Asn	Cys	Tyr	Cys	Lys	Lys	Cys	Cys	Phe

20

25

30

His	Cys	Gln	Val	Cys	Phe	Ile	Thr	Lys	Gly	Leu	Gly	Ile	Ser	Tyr	Gly

35

40

45

Arg	Lys	Lys	Arg	Arg	Gln	Arg	Arg	Arg	Ala	Pro	Pro	Asp	Ser	Glu	Val

50

55

60

His	Gln	Val	Ser	Leu	Pro	Lys	Gln	Pro	Ala	Ser	Gln	Pro	Gln	Gly	Asp

65

70

75

80

Pro	Thr	Gly	Pro	Lys	Glu	Ser	Lys	Lys	Val	Glu	Arg	Glu	Thr	Glu

85

90

95

Thr	Asp	Pro	Val	His											

100

<210> 87

<211> 306

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: tat.SF162.opt

<400> 87

atggagcccg tggacccccc cctggagccc tggaaagcacc cggcagccca gcccaagacc 60
 gcctgcacca actgctactg caagaagtgc tgcttcact gccaagggtgtg cttcatcacc 120
 aaggccctgg gcatcagcta cggccgcaag aagcggccgc agcggccgc cgcccccccc 180
 gacagcgagg tgcacccaggta gagcctgccc aagcggccgc ccagccagcc ccaggccgac 240
 cccaccggcc ccaaggagag caagaagaag gtggagcgcg agaccgagac cgacccctgtg 300
 cactag

306

<210> 88

<211> 306

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
tat.cys22.SF162.opt

<400> 88

atggagcccg tggacccccc cctggagccc tggaagcacc ccggcagcca gcccaagacc 60
gcctggcacca actgctactg caagaagtgc tgcttccact gccaggtgtg cttcatcacc 120
aaggggcttgg gcatacgcta cgccgcgaag aagcgccgcc agcgccgcgc cgcccccccc 180
gacagcgagg tgcaccagg tgcaccgtt gaggcctgccc aagcagcccg ccagccagcc ccagggcgac 240
cccacccggcc ccaaggagag caagaagaag gtggagcgcg agaccgagac cgaccccgta 300
cactag 306

<210> 89

<211> 168

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
tatamino.SF162.opt

<400> 89

atggagcccg tggacccccc cctggagccc tggaagcacc ccggcagcca gcccaagacc 60
gcctgcacca actgctactg caagaagtgc tgcttccact gccaggtgtg cttcatcacc 120
aaggggcttgg gcatacgcta cgccgcgaag aagcgccgcc agcgccgcgc 168

<210> 90

<211> 102

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: tat cys22
SF162 protein

<400> 90

Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser
1 5 10 15

Gln Pro Lys Thr Ala Gly Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe
20 25 30

His Cys Gln Val Cys Phe Ile Thr Lys Gly Leu Gly Ile Ser Tyr Gly
35 40 45

Arg Lys Lys Arg Arg Gln Arg Arg Ala Pro Pro Asp Ser Glu Val
50 55 60

His Gln Val Ser Leu Pro Lys Gln Pro Ala Ser Gln Pro Gln Gly Asp
65 70 75 80

Pro Thr Gly Pro Lys Glu Ser Lys Lys Val Glu Arg Glu Thr Glu
85 90 95

Thr Asp Pro Val His Glx
100

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)